1966-10

An investigation of methods of improving the intelligibility of audio-frequency speech in noise

Huddy, Norman Walter

Monterey, California. U.S. Naval Postgraduate School

http://hdl.handle.net/10945/9554

AN INVESTIGATION OF METHODS OF IMPROVING THE INTELLIGIBILITY OF AUDIO FREQUENCY SPEECH IN NOISE

NORMAN W. HUDDY, JR.

AN INVESTIGATION OF METHODS OF IMPROVING THE INTELLIGIBILITY OF AUDIO-FREQUENCY SPEECH IN NOISE

by

Norman Walter Huddy Jr.
Captain, United States Marine Corps
B.S.E.E., Villanova University, 1959

[redacted]

Submitted in partial fulfillment
for the degree of

MASTER OF SCIENCE IN ENGINEERING ELECTRONICS

from the

UNITED STATES NAVAL POSTGRADUATE SCHOOL
October 1966

1966
Hudd, D.

ABSTRACT

A discussion of the nature of speech is presented, followed by a review of speech processing to date, with emphasis on the characteristics of speech which must be retained for intelligibility. Methods of measuring speech intelligibility are described. The relative merits of abrupt and gradual audio clipping of speech are investigated, and two tone and articulation test results are presented showing that there is no significant difference in these methods of clipping with respect to speech intelligibility. Processing of speech to radio frequencies, filtering and retranslation to audio to improve the peak to average value ratio of the audio frequency prior to transmitting it through a noisy channel is investigated. Two tone and articulation test results are presented showing that this processing results in a 20% improvement in speech intelligibility over audio clipping and filtering.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. Introduction.

In spite of all his attempts to sophisticate his systems of communications, man has yet to devise a more effective means than ordinary speech. While the redundancy and lack of logic of some aspects of speech is obvious, there is no other means available to us that so effectively performs the mission of a communications system, which is to transfer thoughts or ideas from one human brain to another. No other method of communication can so precisely indicate the exact meanings that the individual "transmitting" desires the individual "receiving" to understand. Speech is limited, of course, by language, vocabulary, and so on.

When it is desired, however, to transmit thoughts, or to communicate, over a distance of more than a few feet, we discover that speech has further limitations or drawbacks. When we attempt to use speech in an electronics communications system that is peak-power-limited, and to transmit this speech in a noisy environment, we find that these drawbacks can be serious impediments to effective communications. Hence, for nearly forty years (25) men have been studying ways in which to process speech to aid in achieving better communications. The main idea has been to process speech in certain ways to remove its disadvantages as a communications means, while retaining as much of its ability to convey meaning to the listeners as possible. The measure of the success of a speech processing system has been the degree by which intelligibility is improved, for a given set of conditions, over unprocessed speech. Generally there has not been too much concern, through the years, over obtaining high quality speech reproduction for communications purposes, but only over obtaining high intelligibility.

In the succeeding sections there will be given a brief description of the nature of speech and a review of what types of things have been done in speech processing to date, and with what results. Then there will be a short discussion of methods of determining speech intelligibility, followed by a description of, and comments on the value of two new ideas in speech processing. These ideas consist of the following: First, it might be possible to reduce the distortion introduced by audio speech clipping, which, as we will see, is a common method of speech processing, by choosing a clipper with a gradual input-output characteristic, rather than the normal one wherein clipping occurs abruptly at some particular level. Second, it should be possible to improve the intelligibility of an audio signal by translating it to radio frequencies, (that is generate a single-sideband wave) then clip it, filter and translate it back to the audio range again. The results of intelligibility tests on these systems will be presented and discussed in the hope of providing further understanding of speech and speech processing.

2. The Nature of Speech.

Speech can be compared to a modulated carrier signal (5), the nature of which varies quite a bit with time. For the vowels or voiced sounds, the carrier consists of tones generated by the vocal cords, while for the consonants or unvoiced sounds the carrier is like broadband noise (18). The modulation consists of:

(a) Turning on and off the carrier.

(b) Frequency modulation by emphasis, inflection and so on.

(c) Modification of the harmonic content of the carrier.

(d) Amplitude modulation.

As with any other waveform, speech may be represented in the

frequency domain or the time domain. In the frequency domain we see that for vowels, intensities are concentrated in one or more distinct frequency regions, called formant regions. Each vowel sound has its own set of characteristic formant regions, although these are not necessarily the same when the sound is uttered by different people. The consonants have components in the frequency domain that generally lie higher than those of the vowels and are of lower intensity. Here the intensities tend to be scattered continuously over the spectrum, hence the noise-like description for the carrier of a consonant as given above (10). This distribution of the intensities in consonants is caused by the fact that they are not produced by the vibration of the vocal cords, as are vowels.

The average intensity spectrum of speech is shown in fig. 1 (10). Here we see a sharp drop after about 600 Hz. The formant regions are typically below 3000 Hz. for adult speech and for vowels three are usually found (21). Figure 2 shows the formant regions for the ee sound in "proceedings" where a fourth formant at 4000 Hz is present (21).

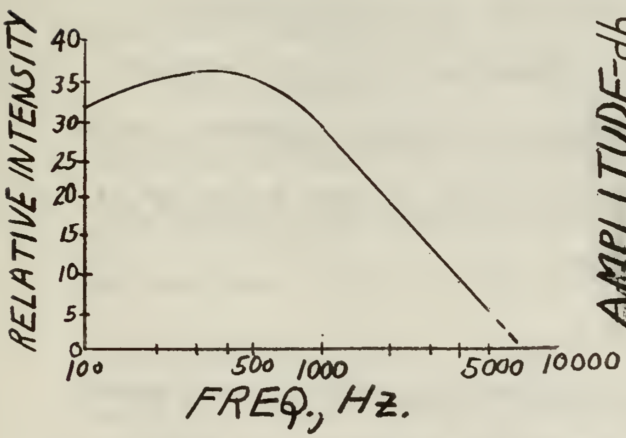


Fig. 1 Intensity distribution of average speech

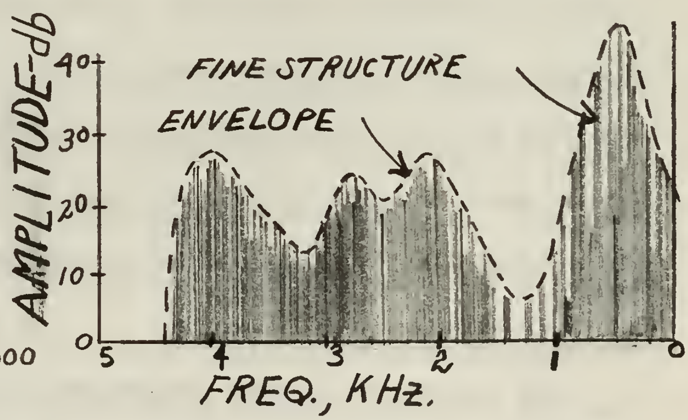


Fig. 2 Spectrum of ee sound in "proceedings".

The formant regions occur at harmonics of the fundamental frequency of the voice which ranges from about 90 Hz. for a deep-voiced man to 300

Hz. for a high-voiced woman (8).

As we will see in our discussion of speech processing, a great deal can be done to speech that will still yield intelligibility. For some time the search has been on to discover what elements in speech remain invariant under these sometimes radical alterations that still result in intelligibility. This search has narrowed down to the frequency spectrum. Agreement has more or less been reached that if the formant regions are not severly altered the intelligibility of the speech will not suffer unduly. The most striking example of this is the formant vocoder. This device locates and measures the energy in the formant regions. This information can be coded, transmitted, and intelligible speech reproduced at the receiver (21). In 1959 here at the U.S. Naval Postgraduate School, S.R. Wilde devised a scheme for speech synthesis using the formant regions that resulted in intelligible speech using only 140 Hz. of bandwidth.

In these vocoders we see that the only information used in the original wave is that contained in the power spectrum. It has been shown that the information contained in the spectrum, the autocorrelation function, and the average number of zero crossings of the time domain waveform are all three equivalent, and that the formant movements can be approximated by the running averages of the number of zero crossings of the original and differentiated waves (2).

3. Speech Processing, General.

The subject of speech processing is generally concerned with answering the following question: What characteristics of speech are undesirable, and what can be done to eliminate them, while not altering the power spectrum of the wave a great deal? In a peak power limited system we are

interested in a signal with a low peak to average value ratio. With such a signal we can achieve the best average signal to average noise ratio when we attempt to transmit our signal through a noisy environment. The normal peak to average ratio of speech, however, is 14.5 db (18). This is an undesirable feature of speech which we would like to eliminate. Also, as we have seen, speech covers a bandwidth of around 5000 Hz. Obviously, it would be nice to reduce this if possible. The following two sections will discuss the efforts that have been put forth to accomplish these two objectives while still retaining intelligibility.

4. Audio Speech Processing.

The first step in the effort to reduce the peak to average ratio of speech was to clip the peaks of the speech wave. In 1946 J.R. Licklider found that for such a system as we have described maximum intelligibility is achieved by clipping the peaks of the speech wave and using the available power for the rest of the wave. He also attempted center clipping wherein the center portions of the wave is removed and only the peaks are passed. This, however, resulted in very poor intelligibility beyond a few db of clipping (12).

The big difference in these two types of clipping is that peak clipping does not alter the zero crossing characteristics of the time wave form while center clipping does. This can be seen in fig. 3. Thus, as we have seen, center clipping alters one of the invariants and we would expect intelligibility to suffer. Licklider also performed various degrees of linear rectification on speech signals and found that articulation began to suffer just as half-wave rectification was reached or just at the point where the zero crossings began to be altered. Figures 4 and 5 show the results Licklider obtained using articulation tests as

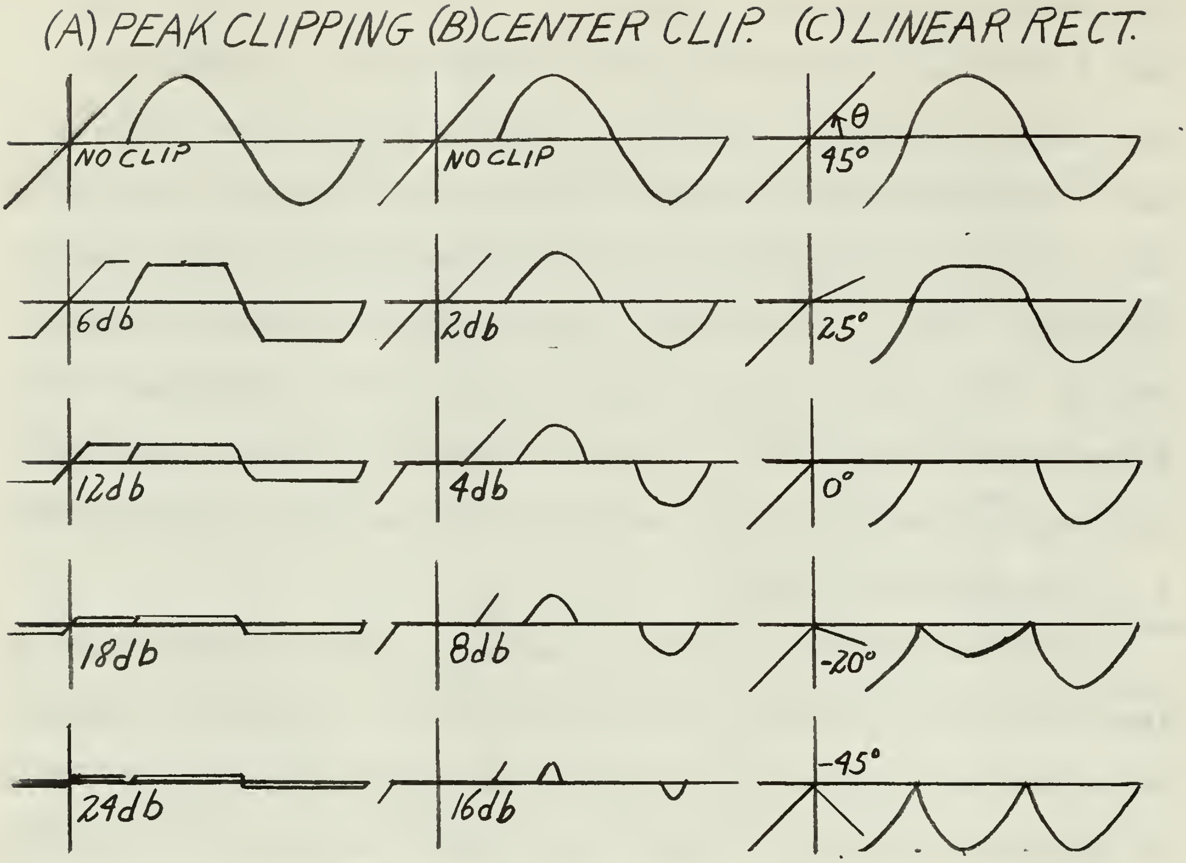


Fig. 3. Characteristics of (A) Peak Clipper
(B) Center Clipper (C) Linear Rectifier

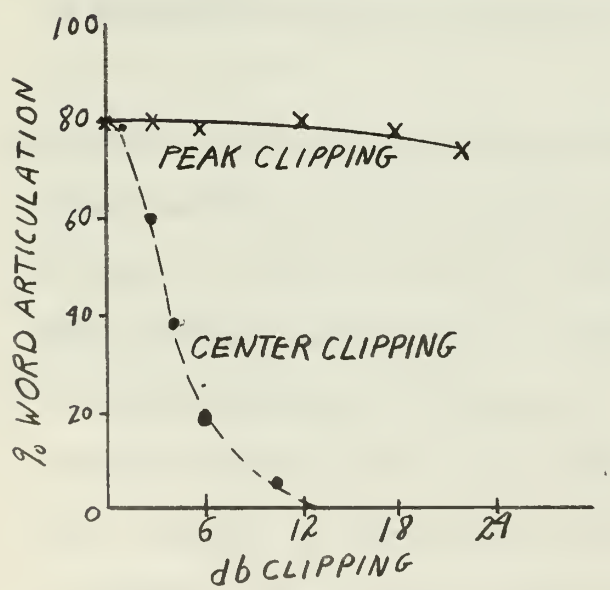


Fig. 4. Effects of peak and center clipping on speech in noise.

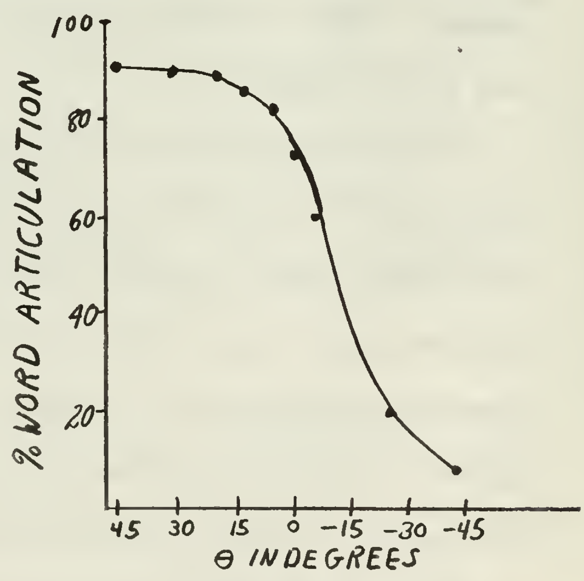


Fig. 5. Effect of linear rectification on speech in noise. θ shown in Fig. 3(C)

the measure of intelligibility.

In 1948, Licklider, together with I. Pollack, applied himself to a further study of the effects of various types of processing on speech intelligibility (13). They investigated the effects of integrating, differentiating, and clipping of the wave form on speech intelligibility without noise. Figure 6 illustrates the effects of various combinations of these steps on a sine wave and a speech wave, as far as appearance in the time domain is concerned. This study discovered the following:

(a) Differentiation and integration alone do not effect intelligibility to a significant degree.

(b) Infinite (very hard) clipping alone causes a decrease of intelligibility of about ten percent below (a).

(c) Infinite clipping preceeded by differentiation caused no significant decrease in intelligibility.

(d) Infinite clipping preceeded by differentiation followed by integration yielded the same results as (c).

(e) Infinite clipping followed by differentiation had no effect on intelligibility other than that caused by the clipping alone, but the quality of the resulting speech was worse.

(f) Infinite clipping followed by integration caused no further degradation of intelligibility over clipping alone, but the quality of the speech was improved.

(g) Infinite clipping preceeded by integration resulted in very poor intelligibility, with scores 70% below those of (a).

(h) Infinite clipping preceeded by integration followed by differentiation resulted in even poorer scores, 80% below those of (a).

The integrator and differentiator used in these tests are shown in

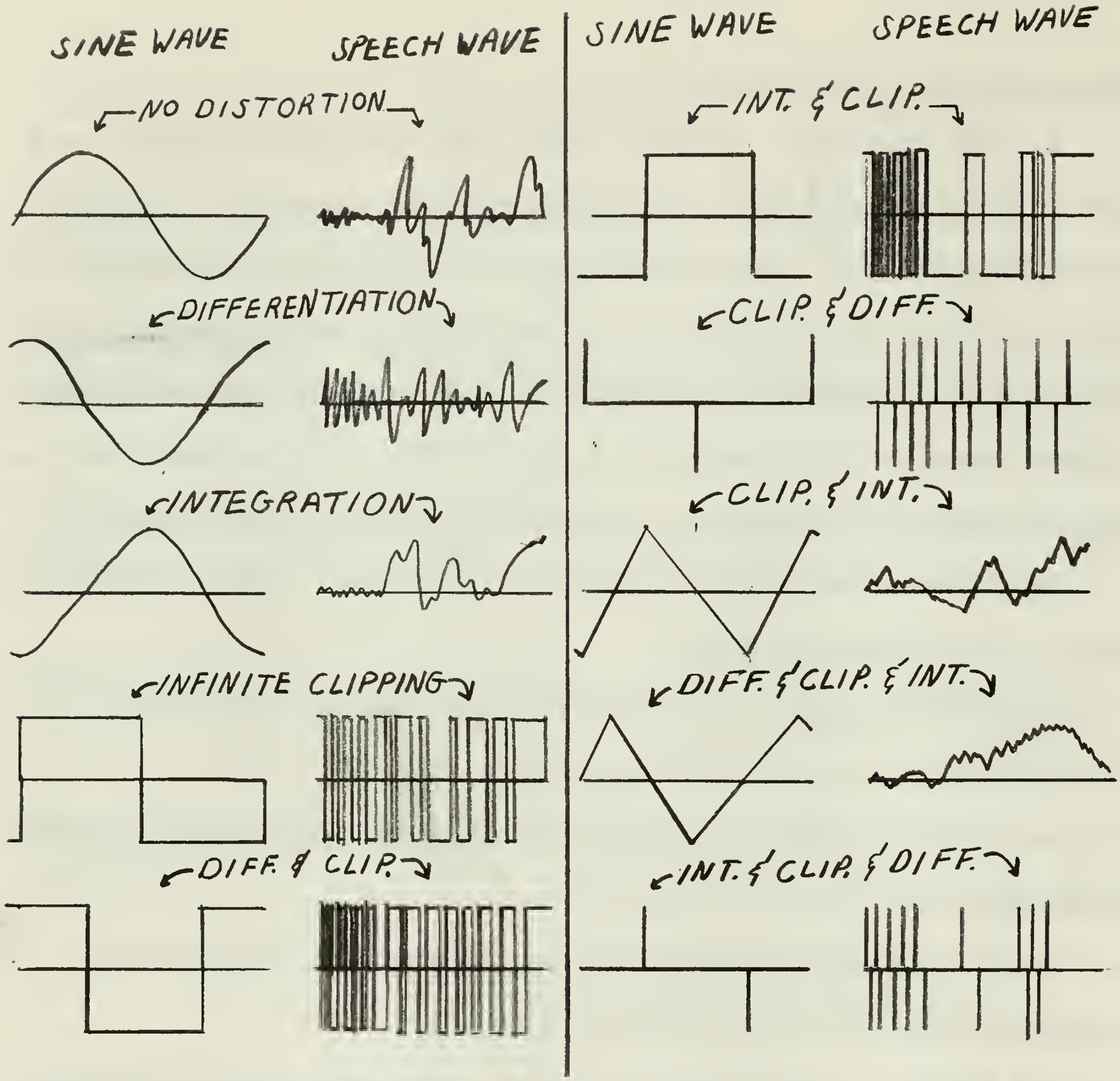


Figure 6. Schematic Illustration of the effects of the distortions upon sine waves and upon speech waves.

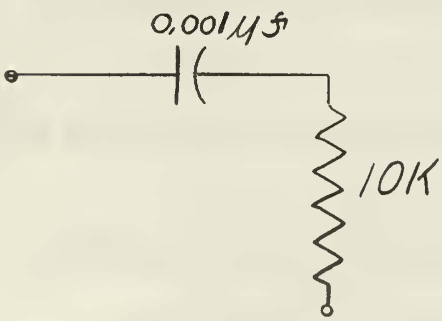


Fig. 7. Differentiator

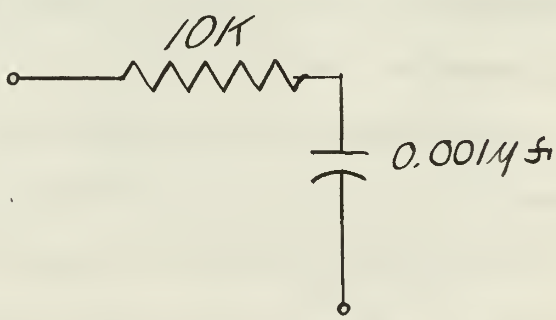


Fig. 8. Integrator

figures 7 and 8. Differentiation serves to "tilt" the spectrum up. It introduces six db less attenuation for each octave increase in frequency. Integration has the opposite effect, tending to tilt the spectrum downward six db per octave. Looking again at fig. 1, we see that the intensity of the high frequencies in speech is much less than that of the low frequencies in natural speech. When we differentiate then clip we are emphasizing the highs before clipping. Thus in the clipped wave, the highs, which carry much of the intelligibility, are less likely to be masked by noise. We are of course changing the quality of the speech in doing this. When we integrate before clipping, we do the opposite and the highs can be completely lost. Since clipping alone tends to bring the lows down closer to the highs in intensity, clipping followed by integration will result in more natural sounding speech. On the other hand clipping followed by differentiation will result in worse speech quality since the normal ratios of intensities is further changed.

Thus we can say that infinite clipping preceeded by differentiation can be used to reduce speech to a bivariate code and integration can be used to retrieve natural speech. However it has been found that differentiation before clipping raises the peak to average ratio of the wave by 4 db to 18.5 db (18). Thus we would have to clip harder and amplify more after clipping. Since we are interested mainly in intelligibility, it is doubtful whether this differentiation is worth it. We can see that integration <u>before</u> clipping is just the opposite of what we want to do with a speech wave.

So far we have discussed clipping only with reference to a fixed signal to noise ratio, or with reference to no noise at all. Pollack discovered that infinite peak clipping improved intelligibility for a

given signal to noise ratio until high signal to noise ratios were reached (19). This decrease in the benefits of clipping is expected since, as we have seen, infinite clipping does reduce intelligibility by about ten percent with no noise present. This can be explained by considering the distortion introduced by clipping as noise. Then, beyond a certain level of actual noise, the noise introduced by clipping will outweigh the benefits gained by clipping (18). In later studies (20) Pollack investigated the effect of clipping on speech further and found that clipping was definitely beneficial at poor signal to noise ratios. For a five db signal to noise ratio he determined that when the peak of the speech wave was clipped 24 db, in order to achieve the same intelligibility the gain had to be increased to 13 db, resulting in an improvement of 11 db.

As has been pointed out previously, it would also be nice if the bandwidth of speech could be reduced. Investigations have been carried out to determine the effects of limiting the frequencies of the speech wave form. Among these were those carried out by Egan and Wiener at the Harvard Psycho-Acoustical Laboratories. These results show that intelligibility scores vary only about eight percent below the full bandwidth case when the speech frequencies are limited to 340 and 3900 Hz. As long as the pass band for speech is in this range intelligibility does not suffer. The important thing is that most of the formant regions must be included in the pass band (7). Figure 9 shows the effect of filtering on the intelligibility of speech.

It has been determined that if speech is limited to a given band of frequencies, the intelligibility of a clipped relative to an unclipped signal is a function of the signal to noise ratio alone (19). We have

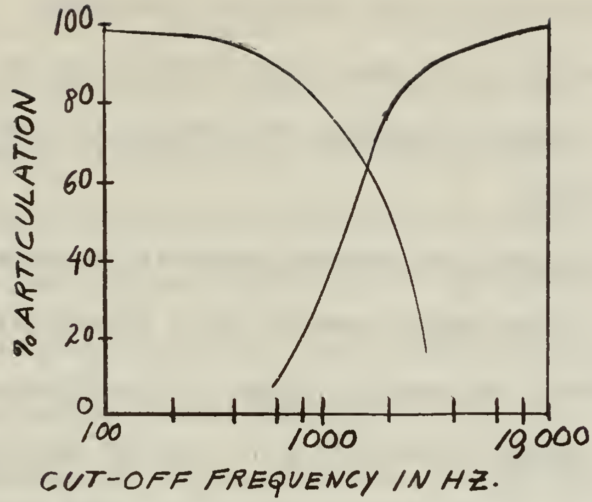


Fig. 9. Intelligibility of band-limited speech.

seen how clipping alone introduces a decrease in intelligibility at high signal to noise ratios. This effect is shown to decrease if the lower frequencies of the speech are removed prior to clipping (19). The lower frequencies contain nearly all voice fundamentals. The formant regions, however, are at harmonics of the voice fundamentals. The clipping process, as we will see in section 8, introduces harmonics of the frequencies contained in the original wave. Thus if the lower frequencies are present when a speech wave is clipped, the harmonics generated by the clipping process lie right where the formant regions should be and thus alter them. Also, as we shall see in section 8, the clipping process introduces intermodulation products among the frequencies present in the original wave. These products will also lie in or near the formant regions if the low frequencies are present in the unclipped wave. Thus we can see that the "noise" generated by clipping can be reduced by removing frequencies below the highest expected voice fundamental, about 300 Hz. We cannot completely eliminate frequency distortion caused by clipping. Harmonics and intermodulation products from all frequencies present in the wave to be clipped will appear as undesired frequency components in the clipped

wave. The thought occurs that perhaps some particular type of clipper can be found that will reduce these undesired components. Section 7 is devoted to a presentation of an idea along these lines and to showing intelligibility test results comparing two divergent types of clippers.

5. R-F Speech Processing.

So far in the discussion of speech processing we have only been considering operations on the speech wave at audio frequencies. In communications systems, however, we usually intend to translate our intelligence to radio frequencies before transmitting it through any appreciable noise. Focusing our attention on radio frequency processing we see that the single sideband system of modulation lends itself very well to a study of such processing. Here we have an opportunity to study the effects of clipping at three places in the system; at the audio frequencies, at r-f, but with the double sideband signal, and at r-f with the single sideband signal. In fact, an extensive study at the Montana State College in 1962 did just that (27). In this project clipping of various degrees was performed at each point in a single sideband system; at audio, double sideband, and single sideband, with appropriate post-clipping filtering to regain bandwidth. The processed signals were mixed with varying degrees of noise and signal intelligibility of the wave after detection was measured. In addition, combinations of clipping at all three places were tested, as well as various methods of achieving high clipping levels, such as clipping one-half the desired amount, filtering, and then clipping the other half.

The results of this study show that single sideband clipping yields significantly higher intelligibility scores than do audio or double sideband clipping, or any combination of the three. When clipping at single

sideband is done, the frequencies being clipped are the same ones as in the original audio wave, but after they have been translated to radio frequencies. Now the formant regions, for instance, no longer bear harmonic relationships to each other. When we clip at r-f, the harmonics and intermodulation products are "splattered" over a much wider frequency range, so it is possible to filter out all but those occurring immediately around the carrier frequency. Thus, when the wave is demondulated we have many fewer undesired components present.

In double sideband clipping we have twice as many frequencies present in the wave to be clipped and so end up with many more undesired components too close to the carrier to filter out without removing our intelligence.

In single sideband clipping we do have a repeaking problem as a result of the filtering. In the Montana study this was observed to reach four db for very hard clipping. However, this is still a considerable saving over the original 17.5 db peak to average ratio of unclipped single sideband speech (18).

If a speech wave is infinitely clipped at the audio level and is used to modulate a single sideband wave with an r-f pass band of f + 300 to f + 3000, the peak to average ratio of the resulting single sideband signal is about 7.3 db (18). Thus, not only do we have more distortion present with audio clipping, but we do not achieve as low a peak to average power ratio as with single sideband clipping.

The effects are so well recognized now that the Collins Radio Company, in their single sideband manual categorically state that speech clipping at audio frequencies "is of no practical value in a single sideband transmitter" (1).

Returning to the Montana study for a moment, this group points out that iterative clipping, that is clip, filter, clip again, has no advantage over single sideband clipping in one stage followed by filtering to regain band width. In addition various combinations of differentiation, integration and clipping were investigated with no significant result (27).

The Voice of America radio has used single sideband clipping to achieve a 9 db improvement in signal to noise ratio in combating jamming (11). Single sideband clipping has been applied to amateur radio also with excellent results (24).

The above discussion of speech processing at radio frequencies was with reference to a system wherein the noise is introduced at the radio frequencies. That is a system which is concerned with transmitting a radio frequency wave through a noisy channel. But consider a peak power limited system where the noise is introduced at the audio frequencies, such as a public address system or the "one MC" and "21 MC" systems aboard U.S. Navy ships. We have seen that it would be advantageous to perform clipping on the audio wave to improve intelligibility. But might it not be feasible to introduce a device into the system in which the signal is translated to a radio frequency, clipped, filtered, then translated back down to the audio frequencies? Should not this process result in even greater intelligibility due to the removal of distortion caused by clipping in the filtering of the clipped wave? This idea will be discussed and investigated in section 9.

6. Intelligibility Measure: The Articulation Test.

We have seen how various types of speech processing used in the past effect speech intelligibility, and we have mentioned two additional ideas that we will discuss further on. But no discussion has been made about

how speech intelligibility is measured.

The most commonly accepted method of testing the intelligibility of a speech processing system is the articulation test. First developed by the Bell Telephone Co., (9) these consist of trained listeners listening to a selected list of sounds, words, or sentences and recording what they hear. The results are compared with the lists actually transmitted through the system under test and a mean articulation score is computed. This is compared against known scores achieved using other systems to determine the relative merits of the system under test with respect to intelligible transmission or reproduction of speech.

There are many ways to conduct articulation tests. The test results shown in the next two sections were obtained using the methods described by the Harvard Pyscho-Acoustical Laboratory study, "Articulation Testing Methods II" (16). In these tests phonetically balanced word lists were used. These are lists in which speech sounds occur with approximately the same frequency as they occur in the English language, and the words are so chosen that there are no very easy or very difficult words in each list. That is, all the words are of uniform, intermediate difficulty. This eliminates "dead wood" words which would always be missed or always be heard correctly and thus give no information on intelligibility.

Word lists rather than sentence lists or sound lists were used for the following reasons: Sound lists require a very careful "talker" and very well trained listeners. Neither were readily available. Sentence lists are easier than word or sound lists in this respect, but the time needed to give and grade tests composed of sentence lists was considered excessive. Twenty phonetically balanced word lists were used. The order of the words on each list was randomized with the aid of a table of ran-

dom numbers and the order of some of these lists were reversed to give additional lists. Care was taken to ensure that the listeners did not hear a list and its inverse version within too short a time, and when it was necessary to use a list for the second or third time care was also taken to make sure a sufficient amount of time had elapsed so that the listeners were not able to recognize the order of the words. A total of 32 lists were generated. Samples of these are given in Appendix II. The Harvard study contains all twenty of the original lists, with the words in alphabetical order.

For each word list the peak list word was determined. This is the word which resulted in the highest amplitude for each list. This word was used to determine the peak signal for each list in order to set the clipping level C and the signal to noise ratio λ, defined below. A list of the peak list words and their relative amplitudes is contained in Appendix II.

These word lists were initially recorded with a signal to noise ratio of 45 db on a Berlant Concertone tape recorder. The microphone used was an Altec 660A dynamic. A peak reading meter on the recorder and a Tektronix 515A oscilloscope was used to keep the recording voice at a constant level.

Both series of tests described in sections 8 and 9 involve clipping and signal to noise ratios. Since we are concerned with random noise and peak power limited systems these parameters were defined as:

$$\lambda = \text{signal to noise ratio} = 20\log_{10} E_s'/E_n$$

$$C = \text{clipping level} = 20\log_{10} E_s/E_c$$

where

E_s' = peak signal at point where noise is introduced

E_s = peak signal after clipping

E_n = r.m.s. noise voltage

The noise voltage was generated by a General Radio Company type 1390-B random noise generator. E_n was measured by a calibrated meter on the face of the generator which was connected directly across its output. E_s, E_s', and E_c were measured with an oscilloscope, using the peak list words.

Each test consisted of two of the phonetically balanced words lists of fifty words each. Each word was given as the last word of a carrier sentence. The carrier sentence used was "The word you should write is __________." Only the word under test, always the last word in the sentence, was recorded by the listener. The carrier sentence was used for two reasons (16). First, the listener is prepared for the test word and the missing of words due to inattention is reduced. Second, the carrier sentence helps to keep the voice level even while recording the lists. A space of three to four seconds between carrier sentences was found to be adequate. As recommended in the Harvard study (31), six listeners were used. In the tests described in section eight these were U.S. Navy enlisted men, all of about 22 years of age. The minimum educational background of this group was three years of college training. Unfortunately this group was not available for the test described in section nine. In these tests 5 listeners were used, three of whom were U.S. military officers and college graduates, one of whom was a U.S. Navy enlisted man with some college training and one of whom was a U.S. Navy enlisted man with a high school education. No significant differences in the scores of these listeners were noted.

To avoid fatigue the testing procedures were as follows: The tests

were grouped into sessions of three tests each, each test being of about 14 minutes duration. Between each test the listeners were given about one minute to adjust headsets, chairs and so on. After each session, which lasted around 45 minutes, a 15 minute break was given. No more than three consecutive sessions were held before stopping for lunch or quitting for the day.

The listening facility was in a small quiet room. Each listening position was numbered and consisted of a chair, a writing space, a volume control and a headset. The headsets were standard 300 ohm communications headsets used by the Navy. To each was added foam earpads to add comfort and to help shield noise.

The listeners recorded what they heard on forms like that shown in Appendix II. In order to ensure that the positions did not effect the scores, the average rank of the scores made at each position was calculated. Similarly to check for significant differences in the listeners, the average rank of each listener's scores was also determined. These two figures were made independent by having the listeners shift positions after each test, thus ensuring that no listener stayed at one position too long. These results, shown in Appendix II, were such that there was no substantial difference in listeners or positions.

All listeners scores are given in Appendix II for each series of tests. Further details on each series of tests may be found in section eight or nine and in Appendix II.

7. Other Intelligibility Measures.

While the articulation test is the most widely accepted method of determining intelligibility, as well as the most obvious, work has been done on other methods as well. These methods are based generally on the

idea that intelligibility is a function of how well the running power spectrum of the wave is preserved by the system under test. In one case (22), equipment was built and tested which compared the running power spectrum of the speech before and after processing and calculated an intelligibility index. This index seemed to compare favorably with articulation test scores. In another case (26), devices were designed to measure the average number of zero crossings of the speech wave. From this information an index of intelligibility was calculated.

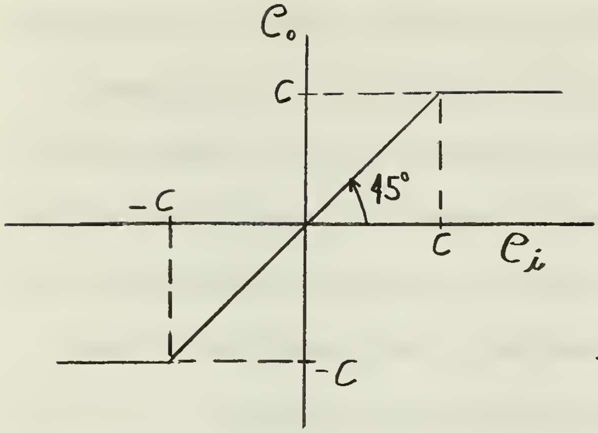
Neither of these two methods seems to have found general acceptance. Hence for this project the more conventional articulation test was used.

8. Gradual and Abrupt Clipping.

As we have seen, speech clipping at audio frequencies can be used as a means to increase the peak to average ratio of speech waveforms in peak power limited systems. We have seen how such clipping can be very beneficial in systems where intelligibility in the presence of noise is of paramount importance, while the quality of the speech heard by the listener is of secondary importance.

Usually one thinks of a clipper as a device having the characteristics shown in figure 10. Here the output e_o is a faithful reproduction of the input e_i up to the point where $e_i = C$. After this point $e_o = C$ no matter how large e_i becomes. This will be referred to as an abrupt clipper, where C is the clipping level.

One can, however, perform clipping with a device with a characteristic such as that shown in fig. 11. Here clipping begins almost as soon as e_i becomes greater than zero and e_o reaches some "saturation" point C, beyond which it remains constant no matter how big e_i becomes. This will be called a gradual clipper.



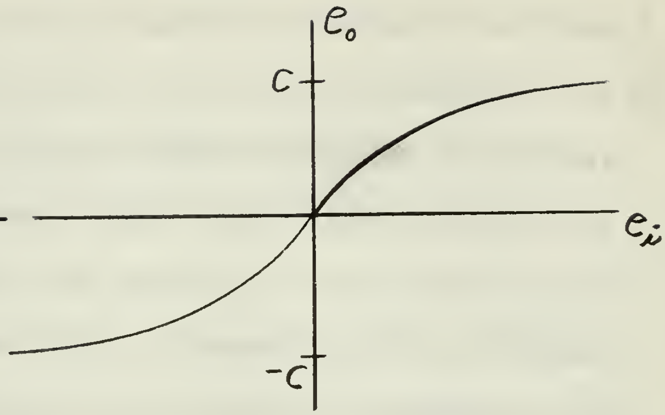


Figure 10. Abrupt Clipper

Figure 11. Gradual Clipper

It is the purpose of this section to investigate the relative merits of the abrupt and gradual clipper as applied to speech. The criteria used will be the intelligibility of the clipped wave in the presence of various degrees of noise with various degrees of clipping.

This investigation was prompted by a remark in an article by Middleton to the effect that gradual clipping has less effect on the spectrum of Gaussian noise than abrupt clipping (15). Davenport has determined experimentally that the probability distribution for the noise-like unvoiced sounds is approximately Gaussian (4), so it would seem that gradual clipping would have some advantage over abrupt clipping.

First it was decided to determine the amount of intermodulation distortion introduced by each type of clipper. In order to do this tests were made on a clipped two tone signal. Tones of 1500 and 2500 Hz. of equal amplitude were combined and clipped by each type of clipper at various clipping levels. The intermodulation components present in the clipped wave were then measured with a wave (spectrum) analyzer.

The gradual clipper consisted of two 1N34A germanium point contact

diodes, arranged back-to-back and unbiased. The clipping characteristics of this device is shown in fig. 12. For the abrupt clipper the same diodes were used, each reverse biased by one volt. The characteristic of this clipper is shown in fig. 13.

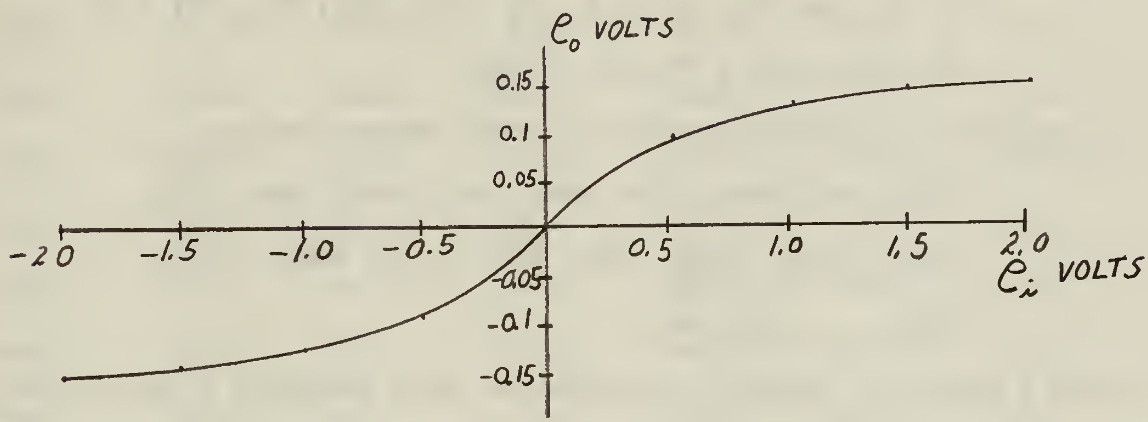


Figure 12. Clipping characteristic, 1N34A, no bias

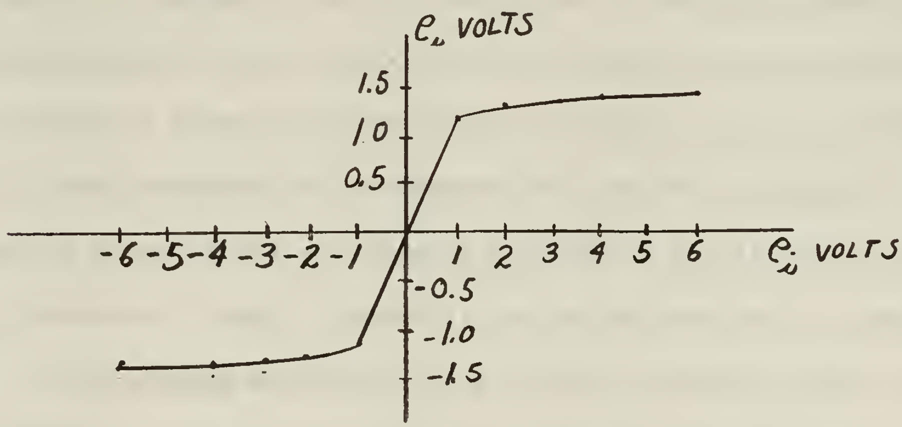


Figure 13. Clipping characteristic, 1N34A, one volt bias

Appendix I shows the equipment setup used in these tests together with a description of the instruments used.

Since the clipper characteristics are odd functions, they can be approximated by an infinite series containing only odd terms, such as:

$$e_o = k_1 e_i + k_3 e_i^3 + k_5 e_i^5 + \ldots$$

Considering only the first five terms of such a series we see that for an input of the form:

$$e_i = A\cos W_1 t + B\cos W_2 t$$

the output will contain the following frequencies (18):

$$W_1,\ W_2,\ 3W_1,\ 3W_2,\ 2W_1 \pm W_2,\ W_1 \pm 2W_2,\ 5W_1,\ 5W_2,\ 4W_1 \pm W_2,$$

$$3W_1 \pm 2W_2,\ 2W_1 \pm 3W_2,\ W_1 \pm 4W_2 \ . \ . \ .$$

For the 1500 and 2500 Hz. tones used these frequencies are:

1500, 2500, 4500, 7500, 5500, 500, 6500, 3500, 7500,

12,500, 8500, 3500, 9500, 500, 10,500, 4500, 11,500,

8500, . . . (all Hz.)

Table I shows the relative amplitudes of these frequency components in db down from the fundamentals when the two tone signal was clipped with the indicated type of clipper. In addition the db difference between the two clippers (gradual minus abrupt) of each component is shown. We assume that we want to retain the two tones in the original signal and that everything else is clipping "noise" which we desire to minimize.

It appears that from the standpoint of intermodulation distortion there is very little difference between the two types of clippers.

Next it was desired to see if either clipper introduced a significantly larger harmonic content when clipping a single tone. A tone of 200 Hz. was chosen to simulate a sound in the range of speech frequencies. Table II shows the results of clipping this tone with each type of clipper.

Here we see that the abrupt clipper does introduce slightly higher harmonic components, especially at the higher frequencies. The difference between the two clippers is small until the higher harmonics are reached. These harmonics, however, are so small that they probably

Clipping level		3.8 db			6.8 db	
Freq.	Gradual clipper	Abrupt Clipper	Gradual - abrupt	Gradual clipper	Abrupt clipper	Gradual - abrupt
500	21.8	19.0	2.8	17.2	13.5	3.7
3500	22.0	20.8	1.2	17.0	14.8	2.2
4500	38.0	28.0	10.0	28.0	27.0	1.0
5500	22.0	21.2	0.8	18.0	14.0	4.0
6500	22.0	21.1	0.9	17.0	14.8	2.2
7500	36.0	28.9	7.1	39.5	29.0	10.5
8500	46.0	54.2	-8.2	40.3	32.0	8.3
9500	38.8	42.2	-3.4	28.5	30.0	-1.5
		17.7 db				
500	13.1	10.5	2.6			
3500	13.4	10.0	3.4			
4500	20.5	16.0	4.5			
5500	13.2	15.0	-1.8			
6500	13.6	10.0	3.6			
7500	31.2	29.8	1.4			
8500	36.1	35.4	0.7			
9500	20.5	21.8	-1.3			

Table I

Distortion components from two tone tests, in db down from fundamental.

Clipping level		6.0 db			12.0 db		
Freq.	Harmonic	Gradual clipper	Abrupt clipper	Gradual - abrupt	Gradual clipper	Abrupt clipper	Gradual - abrupt
200	1st	0	0	0	0	0	0
600	3rd	18.0	13.4	4.6	14.0	11.0	3.0
1000	5th	29.2	27.0	2.2	21.5	18.4	3.1
1400	7th	40.2	42.2	-2.2	27.2	22.8	4.4
1800	9th	51.2	36.2	15.0	32.1	28.2	3.9
2200	11th	72.0	41.2	30.8	46.8	35.6	11.2
2600	13th	72.0	56.2	15.8	41.0	38.6	2.4
3000	15th	*	48.9	-	45.1	39.1	6.0
3400	17th	*	50.0	-	49.2	44.2	5.0
			24.0 db				
200	1st	0	0	0			
600	3rd	12.2	10.2	2.0			
1000	5th	18.0	15.0	3.0			
1400	7th	21.3	18.2	3.1			
1800	9th	24.8	20.8	4.0			
2200	11th	27.1	22.8	4.3			
2600	13th	29.1	24.8	4.3			
3000	15th	31.0	26.2	4.8			
3400	17th	34.0	27.9	6.1			

*Too small to measure

Table II

Single tone clipping results, in db below fundamental

couldn't be detected by the ear. It remains to be seen whether these small differences in intermodulation distortion and harmonic distortion are sufficient to cause a difference in intelligibility, especially if the clipped signal is band limited.

In order to determine whether either clipper results in increased intelligibility it was decided to conduct articulation tests as described in section six. The word lists were played into the clippers at at the levels necessary to obtain the desired clipping levels. The clipped signal was filtered with a pass band of 300 to 3000 Hz. Then noise from the noise generator filtered to the same bandwidth as the speech was introduced at a level corresponding to the desired λ as defined in section six. The clipping levels (C in section six) chosen were 0, 12 db, 24 db, and 33 db. The λ's were 3 db, 6 db, 12 db, and 18 db. The resulting signal was recorded on tape and was played to the listeners later.

Figures 14 (A), (B), (C), and (D) show the results of the articulation tests using the 1N34A's unbiased as the gradual clipper and the 1N34A's with a 1 volt bias as the abrupt clipper. Without the benefit of statistical analysis, one could say that there is very little difference between the two clippers. One might be tempted to say that the abrupt clipper yields slightly higher intelligibility scores than the gradual one. Actually, however, only the sets of points 16 and 17 on fig. 14(B) and 22 and 10 on fig. 14(C) show a statistically significant difference. To determine this, a two-sided Mann-Whitney U test was used. This test is one of the most powerful that can be used on data of this nature (23). The null hypotheses, H_o is that the samples of the two sets of scores being investigated came from the same population. A signifi-

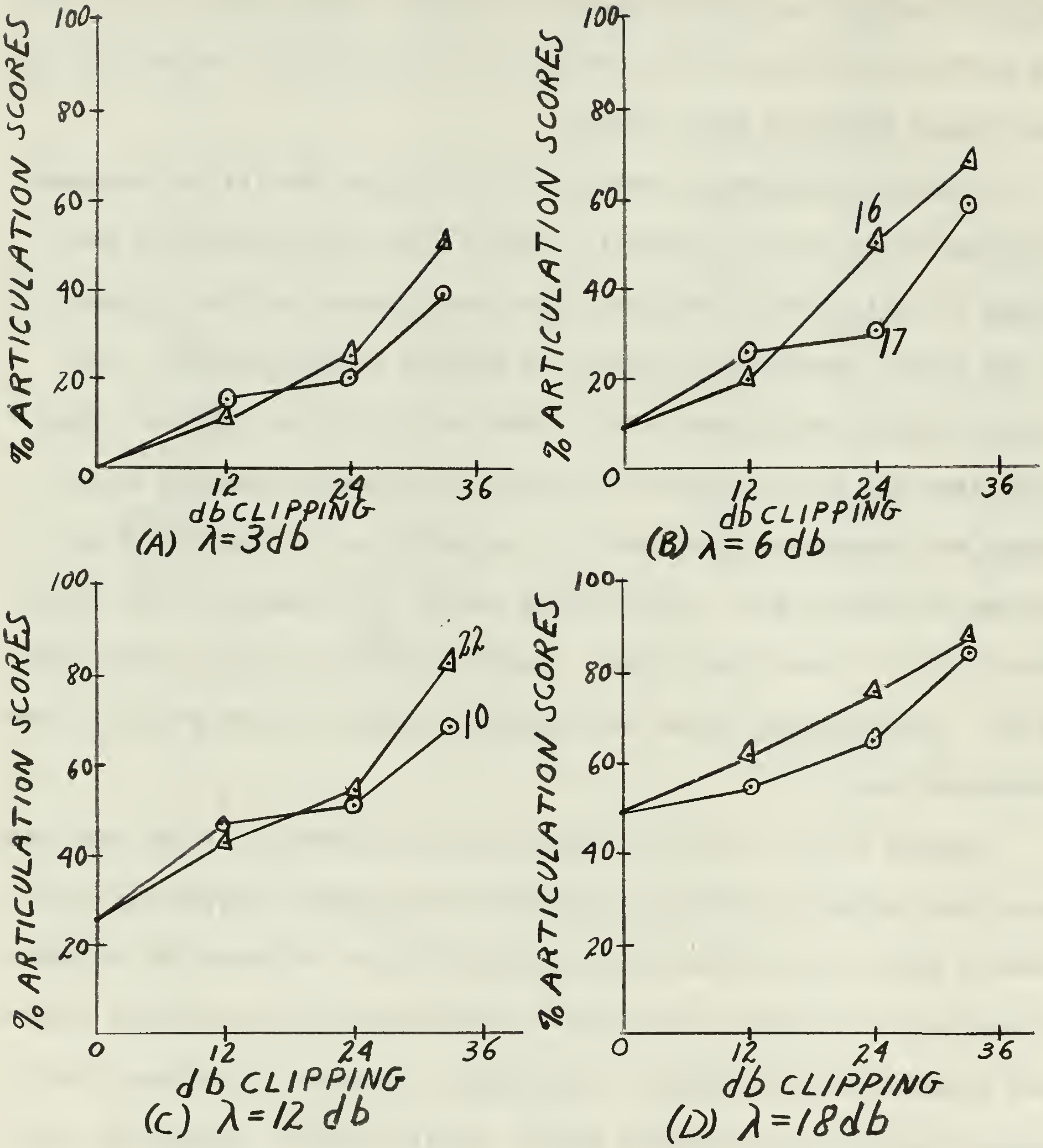


Figure 14. ARTICULATION TEST RESULTS
⊙= 1N34A zero bias, gradual clipper
Δ= 1N34A one volt bias, abrupt clipper

cance level, $\propto$, is chosen and the test determines the probability that the null hypthosis is true. If this probability exceeds the significance level $\propto$, then the null hypothesis is accepted. A significance level of 0.01 was chosen for this data. For a complete description of the Mann-Whitney U Test, with examples, and a discussion of significance levels, see Appendix III.

With the data obtained as described above, and using the Mann-Whitney U Test with a 0.01 significance level, we see that of the twelve sets of data only two caused the null hypothesis to be rejected. Thus it can be concluded that there is no significant difference in the two sets of data, and the fact that the abrupt clipper appears better is just a result of chance.

To confirm this further tests were run using a different gradual clipper composed of two unbiased 1N69A diodes whose clipper characteristic is shown in fig. 15.

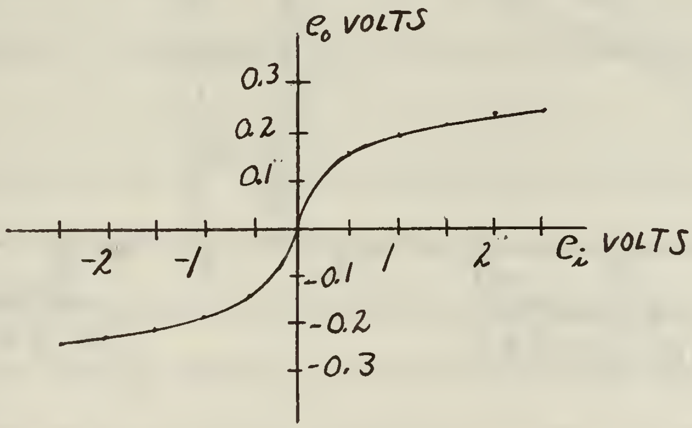


Figure 15. Clipping characteristic 1N69A zero bias

The results of these articulation tests are shown compared with the abrupt clipper results in fig. 16 (A), (B), (C), and (D). Using the Mann-Whitney test again with a significance level of 0.01, again only two sets of scores, marked 28 and 16 and 23 and 35 show significant dif-

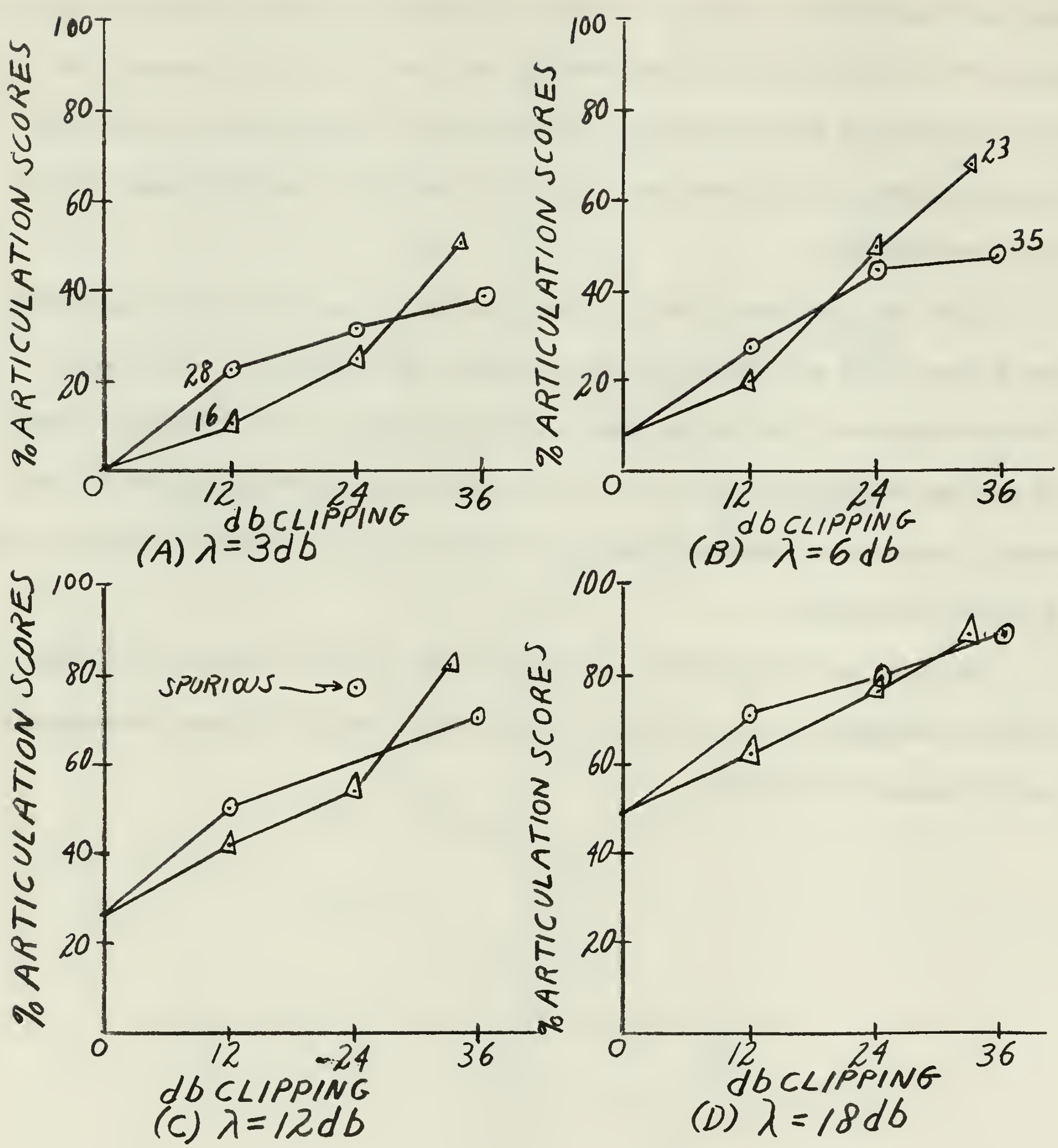


Figure 16. ARTICULATION TEST RESULTS
⊙= 1N69A zero bias, gradual.
△= 1N34A one volt bias, abrupt.

ferences. Note that the differences in these tests are in opposite directions. The point labeled spurious in fig. 16(C) is considered too high and was not used. We see that the abrupt clipper no longer has higher scores.

Thus it is safe to conclude that there is no significant difference between gradual and abrupt clippers as regards speech intelligibility.

In the next section another scheme for speech processing will be considered.

9. Radio Frequency Clipping to Improve Audio Signal Intelligibility.

As we have seen, it is quite well accepted practice to clip a single sideband speech wave in order to improve its peak to average value ratio while retaining intelligibility. This has application in systems in which it is necessary to transmit the radio frequency wave through a noisy channel. In many applications it is desired to transmit speech at audio frequencies through noisy channels. As mentioned before, examples of peak power limited systems in which this is done are ordinary public address systems.

We have also seen that it would be advantageous to perform clipping on the audio wave directly to improve intelligibility. But we have noted that a great number of harmonic and intermodulation distortion components are formed by this clipping process. To reduce this distortion we can translate the audio wave to a radio frequency, say as an upper sideband signal, clip it then filter it to regain the original upper sideband bandwidth. The distortion components introduced by the clipping are now separated by frequencies of the order of magnitude of the carrier, with the exception of the lowest order terms. Passing the clipped r-f wave through a filter such as an upper sideband mechanical filter with a pass

band of the order of magnitude of the audio range, say 3 KHz., will eliminate all but the desired audio and these lowest order distortion terms. It remains to be seen whether the amount of repeaking involved in the filtering and frequency translation of the clipped wave back down to audio cancels out the gain in intelligibility due to the reduction in distortion.

To determine the validity of the above statements, a device which will be referred to as an "R-F Speech Processer" was constructed. A block diagram of this device is shown in figure 17, and detailed diagrams of each component are contained in Appendix IV. The audio input signal is translated to a double sideband signal by the balanced modulator, using the 455 KHz. L-C oscillator to provide the carrier. The lower sideband is removed by the first upper sideband filter. The signal is then amplified and clipped by the r-f amplifier-clipper. This signal is filtered by the second upper sideband filter and returned to audio by the product detector, again using the 455 KHz. L-C oscillator to insert the carrier.

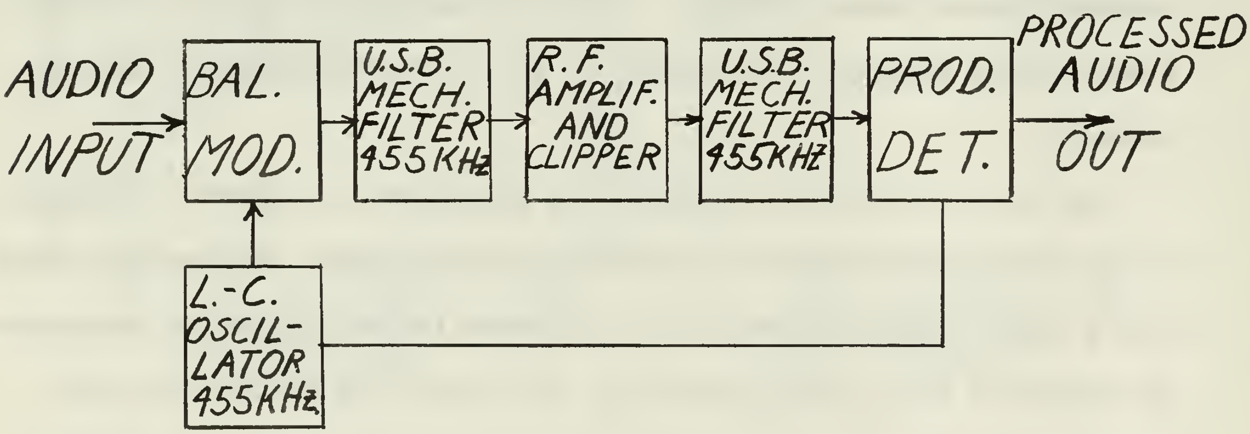


Figure 17. R-F Speech Processer

To compare the intermodulation distortion generated by the speech processer, and to measure the repeaking involved in the filtering and frequency translation of the clipped wave, two-tone tests were used.

The same two tones used in section 8, 1500 and 2500 Hz. were used here. Table III shows the results of these tests, in the r-f column, while the results of the audio clipping with the gradual clipper from section 8 are shown for comparison in the a-f column. In Table III we see quite distinctly that the R-F Speech Processer causes considerably less intermodulation distortion than the audio clipping. The results of the repeaking measurements are shown in Table IV. Here we see that no serious repeaking occurs in the filtering and tranlation of the clipped r-f wave to audio frequencies. (The repeaking of a 20db clipped audio wave filtered from 300-3000 Hz. is 4.2db (28)).

In order to determine the effect of this processing on intelligibility, it was decided to conduct articulation tests with speech processed in this manner. Using the notation introduced previously, 10 tests were conducted, with r-f clipping levels of 12 and 24 db and λ's of 3, 6, 12, and 18 db and the maximum obtainable λ with each clipping level. Because of the small number of tests involved the pre-recorded method of testing was not used. Instead, the word lists described in section 6 were played through the speech processer directly into the listener's headsets for each condition described above. Further details on the equipment setup used in these tests are given in Appendix II.

The results of these tests are shown in figure 18. Further details on test results may also be found in Appendix II. In figure 18 we have taken the average of the three audio clipping test scores obtained in section 8 for each condition shown and plotted them on the same axes as the r-f clipping articulation scores. It can be seen that in each case the r-f clipped speech is more intelligible than the speech clipped and filtered at audio.

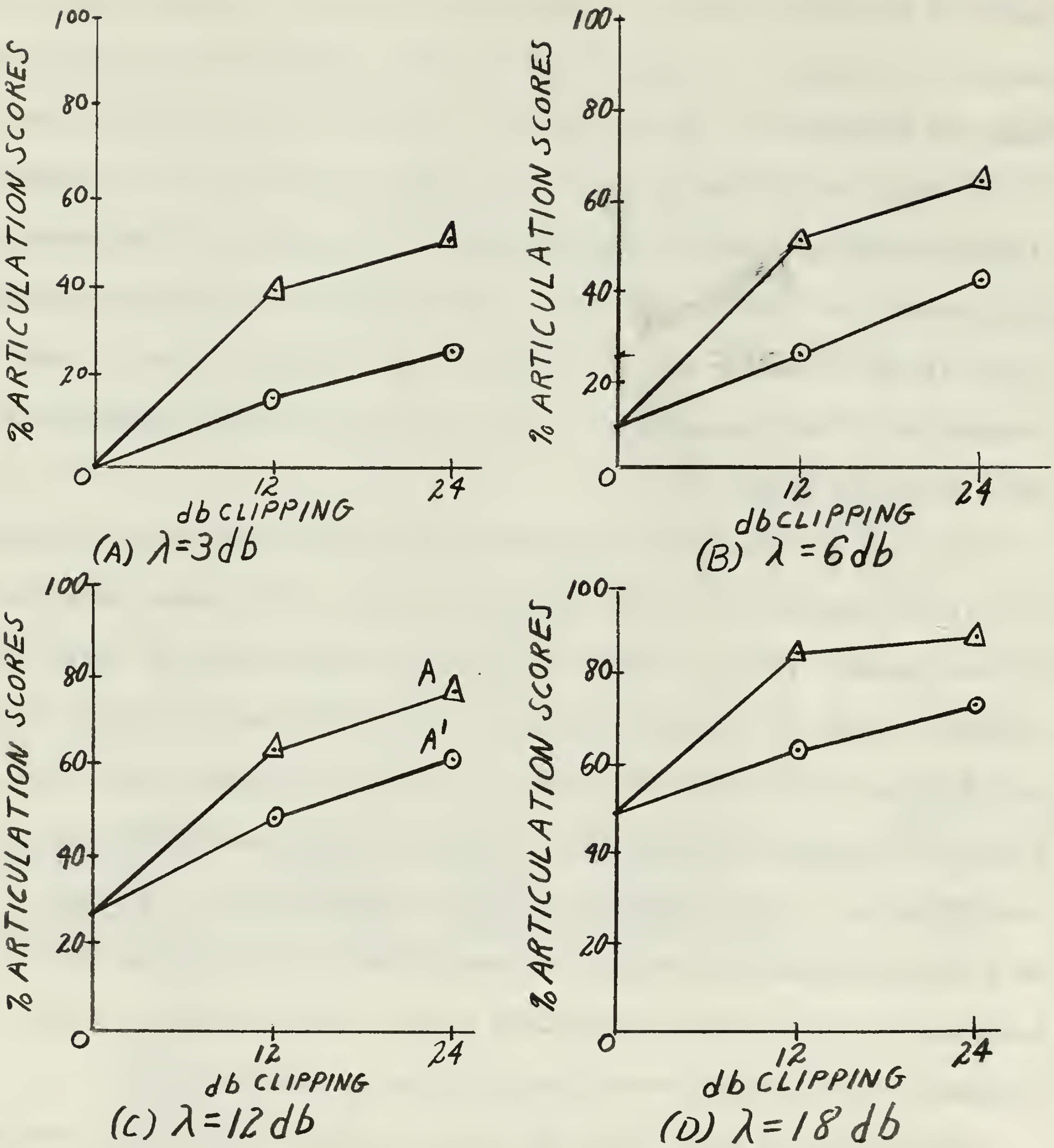


Figure 18. ARTICULATION TEST RESULTS
⊙= Average scores, audio clipping.
△= R-F clipping.

Clipping Level	3.8 db		6.8 db		17.7 db	
Frequency	a-f	r-f	a-f	r-f	a-f	r-f
500	21.8	60.0	17.2	47.1	13.1	40.5
3500	22.0	37.5	17.0	28.5	13.4	21.5
4500	38.0	61.0	28.0	57.2	20.5	50.5
5500	22.0	62.0	18.0	59.0	13.2	52.5
6500	22.0	55.0	17.0	50.4	13.6	47.0
7500	36.0	57.0	39.5	52.5	31.2	52.8
8500	46.0	-	40.3	-	36.1	61.1
9500	38.8	-	28.5	-	20.5	-

Table III. Intermodulation distortion of two tones of 1500 and 2500 Hz. by r-f and a-f clipping, in db down from fundamentals.

Clipping Level db	Repeaking db
3.8	0.6
6.8	1.6
17.7	5.1
26.3	6.2

Table IV. Repeaking associated with filtering and translation to audio of r-f wave clipped to levels shown.

Using the highest set of the three audio clipping scores and applying the U-test, again at a level of 0.01, we find that only two sets of points (r-f and a-f) do not show a statistically significant difference. These are marked A and A' on figure 18c. Thus we can conclude that processing speech with our r-f speech processer is indeed advantageous. The average improvement in articulation over the audio processing is 20.5%.

In addition to the above, tests were run at the best λ available through the processer to determine the effect of the r-f processer alone on intelligibility. At 12 db of clipping the best λ obtainable was 36.5 db, while at 24 db of clipping the best λ was 30 db. The articulation scores obtained under these conditions were 93% for the 12 db case and 90% for the 24 db case. When we consider that even under the best conditions a few words will be missed by the best listener, we can realize that these scores really indicate that r-f clipping and filtering alone have an almost negligible effect on intelligibility. In fact the loss of intelligibility that did occur could be attributed to distortion introduced in the balanced modulator or product detector and might be independent of the actual clipping and filtering process.

10. Conclusions.

We have seen that as long as the formant regions are not too severly distorted or the zero crossings of the time waveform are not radically altered, we can do a lot to speech to improve its characteristics vis-a-vis our communications systems while not impairing its intelligibility. We have noted that this is due to the natural redundancy of speech.

In our investigation of clipping we have discussed the "noise" introduced by the clipping process itself. We have discussed and investigated two ideas for the minimization of this noise. One of these, the idea of

gradual versus abrupt clipping, we found to be of no practical value, except that we now know that if we are given a choice we might as well avoid the need for biasing and use a gradual, unbiased diode clipper rather than an abrupt, biased one, since they will result in the same level of intelligibility. The other idea, of processing the speech at r-f, shows merit. We found that an increase of 20% in intelligibility could be achieved over ordinary audio clipping by this method. This confirms the ideas about the "noise" introduced by clipping and shows how it is reduced substantially by the filtering of the clipped r-f wave.

11. Acknowledgements

The author would like to express his sincere appreciation to the Personnel and Electronics Departments of the U.S. Naval Postgraduate School for their cooperation in obtaining listeners for the articulation tests described, and to fellow students Capt. J.J. Stewart, USMC, Lt. J.W. Lillis, and Lt. W.D. McKay, both USN, for donating their time to act as listeners for these tests. He would also like to thank Dr. Gerald D. Ewing for his constant guidance and infinite patience.

BIBLIOGRAPHY

1. Collins Radio Company. Fundamentals of Single Sideband. Cedar Rapids, Iowa: Collins Radio Company, 1959.

2. Chang, S., Pihl, G.E., and Essigmann, M.W., "Representation of Speech and Sounds and Some of Their Statistical Properties", Proceedings of the Institute of Radio Engineers, Vol. 39, (February, 1951), 147-153.

3. Cherry, Colin. On Human Communication. New York: John Wiley and Sons, 1957.

4. Davenport, W.B., "An Experimental Study of Speech Wave Probability Distributions", Journal of the Acoustical Society of America, Vol. 24, (July, 1952), 390-399.

5. Dudley, H., "The Carrier Nature of Speech", Bell System Technical Journal, Vol. 19, (19 October 1940), 495-515.

6. Dukes, J.M.C., "The Effect of Severe Amplitude Limitation on Certain Types of Random Signal; A Clue to the Intelligibility of 'Infinitely' Clipped Speech", Proceedings of the Institute of Electrical Engineers (London), Vol. 102-103, Pt. C, 1955-56, 88-97.

7. Egan, J.P. and Wiener, "On the Intelligibility of Bands of Speech in Noise", Journal of the Acoustical Society of America, Vol. 18, 1946, 435-441.

8. Fletcher, H., Speech and Hearing in Communication. New York: D. Van Nostrand Company, 1953.

9. Fletcher, H., and Steinberg, J., "Articulation Testing Methods", Bell System Technical Journal, Vol. 8, (8 October 1929) 806-854.

10. French, N., and Steinberg, J., "Intelligibility of Speech Sounds", Journal of the Acoustical Society of America, Vol. 19, 1947, 90-119.

11. Kahn, L., "The Use of Speech Clipping in Single Sideband Communications Systems", Proceedings of the Institute of Radio Engineers, Vol. 45, Pt. 3, (August, 1957), 1148-1149.

12. Licklider, J., "Amplitude Distortion and Speech Intelligibility", Journal of the Acoustical Society of America, Vol. 18, 1946, 429-434.

13. Licklider, J., and Pollack, I., "Effects of Differentiation, Integration, and Infinite Peak Clipping on the Intelligibility of Speech", Journal of the Acoustical Society of America, Vol. 20, 1948, 42-51.

14. Lindgren, B., Statistical Theory, New York: The Macmillan Company, 1962.

15. Middleton, D., and Van Vleck, J., "The Spectrum of Clipped Noise", Proceedings of the Institute of Electrical and Electronics Engineers, Vol. 54, (January, 1966), 2-19.

16. Office of Scientific Research and Development, Articulation Testing Methods II, OSRD Report No. 3802, Harvard University, 1944.

17. Office of Scientific Research and Development, The Effects of Amplitude Distortion upon the Intelligibility of Speech, OSRD Report No. 4217, Harvard University, 1944.

18. Pappenfuss, E., Bruene, W., and Schoenike, E., Single Sideband Principles and Circuits. New York: McGraw-Hill Book Company, 1964.

19. Pollack, I., "On the Effect of Frequency and Amplitude Distortion on the Intelligibility of Speech in Noise", Journal of the Acoustical Society of America, Vol. 24, 1952, pp 538-540.

20. Pollack, I., and Pickett, J., "Intelligibility of Peak-Clipped Speech at High Noise Levels", Journal of the Acoustical Society of America, Vol. 31, 1959, 14-16.

21. Schroeder, M., "Vocoders: Analysis and Synthesis of Speech", Proceedings of the Institute of Electrical and Electronics Engineers, Vol. 54, (May, 1966), 720-733.

22. Schwarzlander, H., "Intelligibility Evaluation of Voice Communications", Electronics, (May 29, 1959), 88-91.

23. Siegel, S., Non-Parametric Statistics. New York: McGraw-Hill Book Company, 1956.

24. Squires, W. and Clegg, E., "Speech Clipping for Single Sideband", Q.S.T., July, 1964, 11-15.

25. Steinberg, J., "Effects of Distortion upon the Recognition of Speech Sounds", Journal of the Acoustical Society of America, Vol. 1, 1929, 121-136.

26. Stuckey, C., and Robertson, D.,"Investigation of an Automatic Voice Intelligibility Concept", Institute of Electrical and Electronics Engineers Transactions on Audio, Vol. AU-12, (January-February, 1964), 3-8.

27. U.S. Air Force Cambridge Research Laboratories. Speech Signal Processing and Applications to Single Side-Band. AFCRL 62-64, Montana State College, 1962.

28. U.S. Army Signal Corps. Premodulation and Postmodulation Clipping in Single Sideband Transmission. AEPG-SIG. 960-86, University of Arizona, 1960.

29. Wathen-Dunn, W., and Lipke, P. "On the Power Gained by Clipping Speech in the Audio Band," Journal of the Acoustical Society of America, Vol. 30, Pt. 1, 1958, 36-40.

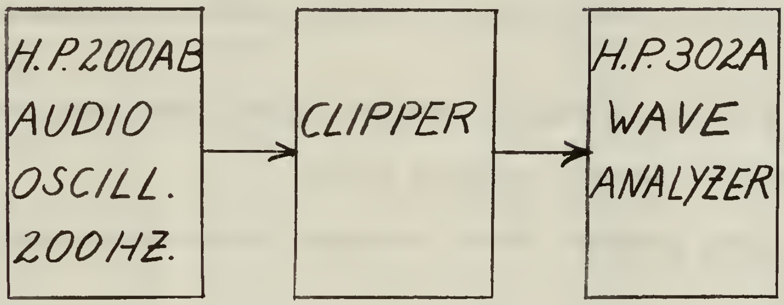
30. Wilde, J., "A Speech Analysis and Compression Scheme for Bandwidth Reduction,"Masters Thesis, U.S. Naval Postgraduate School, 1959.

APPENDIX I

SINGLE AND TWO TONE TESTS

1. Single tone tests.

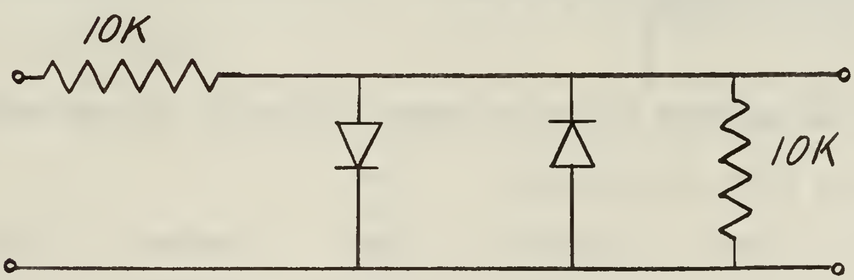
The equipment arrangement used for the single tone tests is shown below:



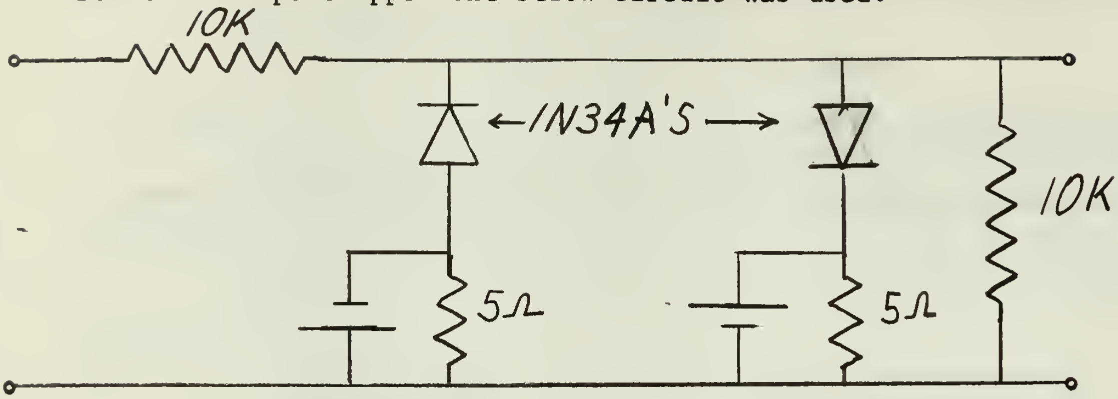
The Hewlitt Packard 200AB audio oscillator, when set at 200 Hz. provided the following output:

Frequency	db
200	0
400	-56.0
600	-67.0

The clipper chassis for the gradual clipper is shown below:



For the abrupt clipper the below circuit was used.

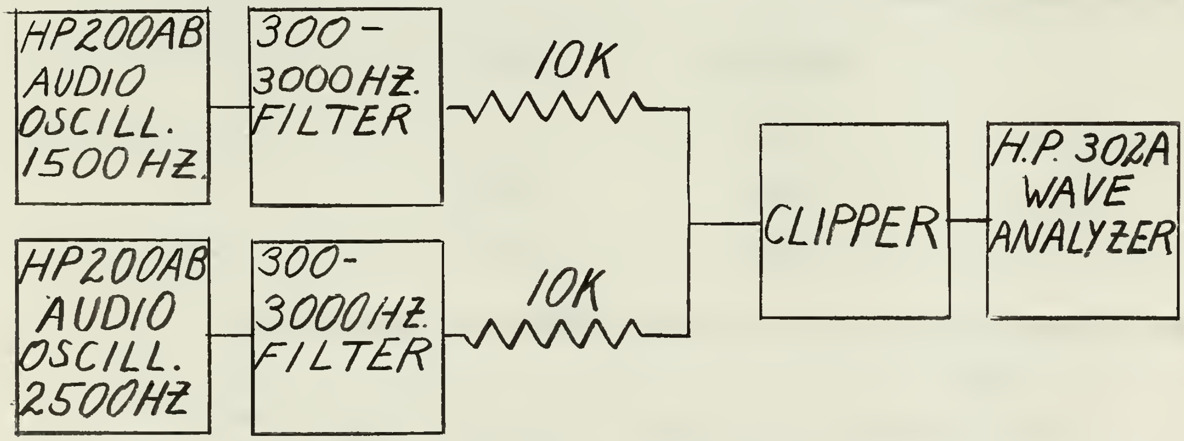


The bias was provided by Hewlitt Packard 721A's, which were adjusted to give a symmetrical one volt clipping level.

The H.P. wave analyzer has an accuracy of 1% + 5 cps and ± 5% in voltage.

2. Two tone tests.

The equipment arrangement used here was as shown below:



The output of the two tone generator, taken at point A, with no clipper attached, across a 10k ohm load was:

Frequency	db	Frequency	db	Frequency	db
1500	0	3500	-69.0	6500	-72.0
2500	0	4500	-71.1	7500	-60.2
3000	-68.0	5000	-72.0		

The clippers used in these tests were identical with those used in the single tone tests.

APPENDIX II

ARTICULATION TEST DETAILS

1. Word lists.

Below are shown two examples of the phonetically balanced word lists used in the articulation tests.

Word List #17

1.	flag	26.	read
2.	thank	27.	year
3.	chess	28.	lit
4.	club	29.	hoof
5.	phone	30.	smart
6.	odd	31.	give
7.	birth	32.	cud
8.	carve	33.	mass
9.	boost	34.	root
10.	grace	35.	throne
11.	foe	36.	ditch
12.	weak	37.	wipe
13.	arch	38.	clown
14.	gate	39.	sip
15.	itch	40.	wild
16.	crowd	41.	spud
17.	troop	42.	ice
18.	beef	43.	key
19.	nerve	44.	toad
20.	with	45.	noose
21.	fume	46.	rude
22.	bit	47.	pact
23.	fuse	48.	than
24.	ten	49.	fluff
25.	nuts	50.	chest

Word List #32

1.	fast	26.	rouge
2.	soak	27.	wise
3.	clog	28.	pad
4.	did	29.	judge
5.	roast	30.	sigh
6.	retch	31.	in
7.	beard	32.	eye
8.	click	33.	pew
9.	cart	34.	rout (rowt)
10.	joke	35.	souse
11.	gang	36.	fair
12.	tilt	37.	wash
13.	ace	38.	crate
14.	hump	39.	seed
15.	mow (mō)	40.	walk
16.	bare	41.	skid
17.	duke	42.	lid
18.	through	43.	pack
19.	puss	44.	theme
20.	web	45.	quip
21.	get	46.	salve
22.	brass	47.	robe
23.	gob	48.	slush
24.	slice	49.	flash
25.	ramp	50.	cork

2. Peak list words.

Word List	Peak list word	Relative amplitude
1.	bask	1.0
2.	perk	1.0
3.	fern	1.1
4.	start	1.05
5.	thrash	1.15
6.	check	1.2
7.	rack	1.0
8.	cloak	1.1
9.	good	1.05
10.	thud	1.1
11.	kept	0.95
12.	kept	1.0
13.	scout	0.90
14.	dope	1.0

Word List	Peak list word	Relative Amplitude
15.	dumb	0.8
16.	look	1.0
17.	ditch	1.0
18.	ditch	1.0
19.	lap	0.8
20.	put	1.0
21.	dull	0.9
22.	crutch	1.0
23.	out	1.2
24.	foot	1.15
25.	soap	0.95
26.	dead	0.85
27.	wreck	0.85
28.	tire	0.70
29.	shock	0.85
30.	thorn	1.0
31.	gyp	0.80
32.	route	1.0

3. Listener's average rank on all tests.

	Section 8 Tests						Section 9 Tests				
Listener	A	B	C	D	E	F	A	B	C	D	E
Ave. Rank	3.9	2.4	3.9	2.5	2.5	3.3	2.1	3.2	2.9	2.5	4.2

4. Position's average rank on all tests.

	Section 8 Tests						Section 9
Position	1	2	3	4	5	6	Not done
Ave. Rank	3.0	2.2	3.3	2.9	3.7	3.6	

5. The tests were given in random order. The below list shows the order in which the tests were given.

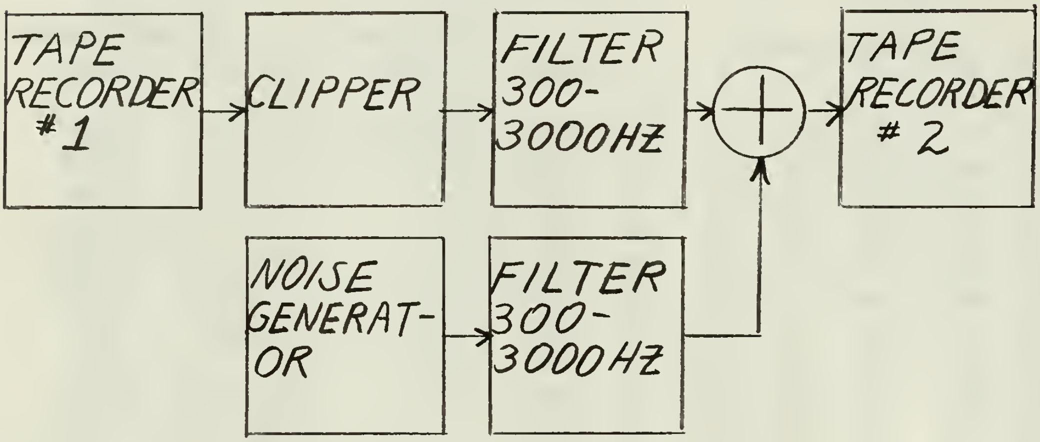
Section 8				Section 9
37	39	25	41	2
38	10	9	32	1
35	24	19	30	3
22	27	40	12	4
18	4	29	31	9
17	7	14	42	10
1	6	15		5
2	36	3		7
16	34	11		6
13	5	33		8
21	20	28		11
12	23	26		12

6. The table below shows further details of the articulation testing described in section eight. The clipper designation 1N34A/0 means the 1N34A diode with zero bias. The 1N34A/1 means the 1N34A diode with one volt reverse bias.

Test #	Listener Scores A	B	C	D	E	F	Ave.	Clipper Used	C db	λ db	Word Lists Used
1	50	55	53	50	67	57	55	1N34A/0	12	18	25,26
2	42	44	34	43	43	42	41	"	"	12	27,28
3	21	29	27	24	26	27	26	"	"	6	1,9
4	14	12	12	22	18	14	15	"	"	3	7,9
5	53	69	67	65	69	68	65	"	24	18	12,14
6	48	47	54	55	57	47	51	"	"	12	3,5
7	27	30	28	33	28	34	30	"	"	6	2,7
8	18	18	16	20	20	25	20	"	"	3	6,8
9	82	88	85	88	82	84	85	"	33	18	15,21
10	66	68	63	75	70	63	68	"	"	12	14,16
11	58	62	54	61	63	55	59	"	"	6	2,6
12	36	43	21	40	53	37	39	"	"	3	22,26
13	61	63	59	71	53	67	62	1N34A/1	12	18	31,32
14	51	38	42	42	43	33	42	"	"	12	30,31
15	17	23	10	32	23	17	20	"	"	6	1,10
16	10	5	10	16	17	15	11	"	"	3	29,30
17	73	81	67	81	83	72	76	"	24	18	21,23
18	56	62	53	57	54	39	54	"	"	12	17,19
19	49	64	40	44	51	49	50	"	"	6	20,22
20	19	28	25	34	34	21	25	"	"	3	15,18
21	89	88	85	89	92	91	89	"	33	18	3,4
22	70	83	78	91	84	84	92	"	"	12	13,15
23	55	75	75	69	72	61	68	"	"	6	11,19
24	58	51	49	56	54	35	51	"	"	3	18,20
25	67	78	71	62	71	79	71	1N69A/0	12	18	13,17
26	57	55	45	49	52	49	50	"	"	12	16,13
27	20	28	23	45	23	30	28	"	"	6	22,24
28	21	24	22	31	28	22	23	"	"	3	11,18
29	86	82	70	80	79	77	79	"	24	18	29,27
30	70	78	67	81	81	82	77	"	"	12	27,21
31	47	66	38	57	60	57	45	"	"	6	24,30
32	26	30	30	37	40	29	32	"	"	3	14,17
33	83	90	88	94	87	90	89	"	36	18	4,8
34	59	80	66	75	71	71	70	"	"	12	8,10
35	45	36	37	44	48	48	47	"	"	6	29,32
36	32	43	33	50	37	40	39	"	"	3	4,6
37	53	60	55	61	61	63	59	NONE	0	18	30,31
38	25	26	24	36	29	25	28	"	"	12	18,19
39	16	18	17	9	19	10	15	"	"	6	1,2
40	40	56	35	52	54	56	49	"	"	18	4,8
41	25	25	23	27	28	25	26	"	"	12	29,30
42	6	10	4	9	11	6	9	"	"	6	6,10

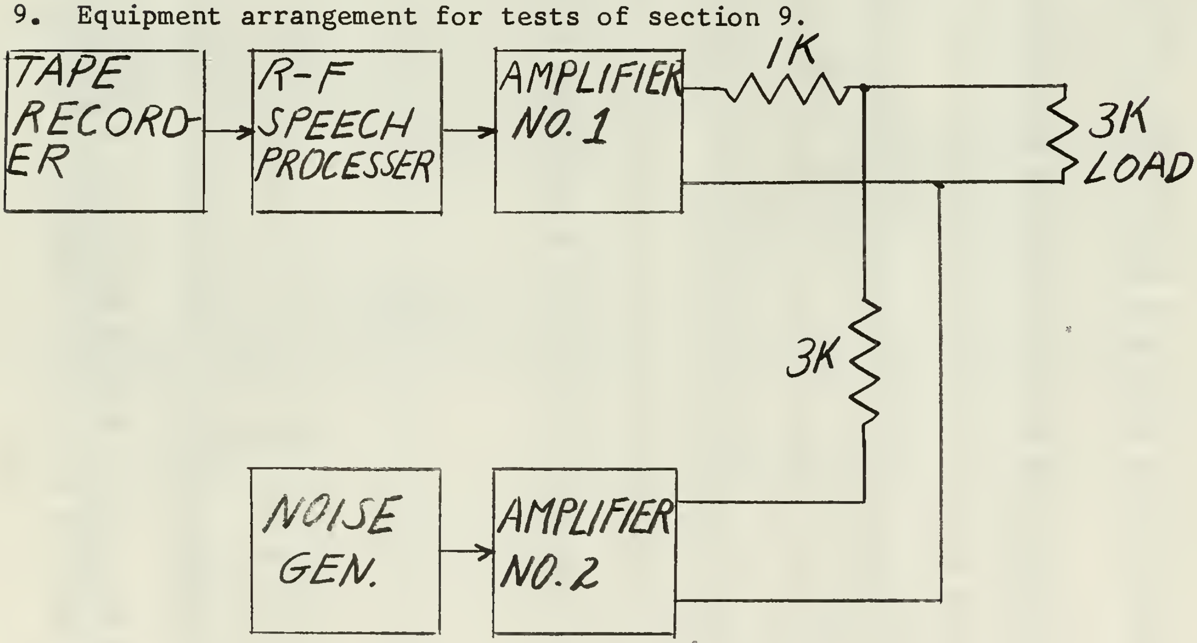
Tests 37, 38, and 39 were composed of speech with no processing at all and were used as the dummy training tests. Tests 40, 41, and 42 consisted of unclipped but filtered speech.

7. Equipment set up for recording tests of section eight.



8. The form shown on the next page was used for all listening tests.

9. Equipment arrangement for tests of section 9.



Amplifier #1 was the amplifier section of an ME-6D/U multimeter, with a flat response from 15 to 250,000 Hz. Amplifier #2 was a Hewlett-Packard 450A amplifier with a flat response from 5 Hz. to 1 MHz.

NAME ______________________ TEST # ________ POSITION # ________

1	____________	26	____________
2	____________	27	____________
3	____________	28	____________
4	____________	29	____________
5	____________	30	____________
6	____________	31	____________
7	____________	32	____________
8	____________	33	____________
9	____________	34	____________
10	____________	35	____________
11	____________	36	____________
12	____________	37	____________
13	____________	38	____________
14	____________	39	____________
15	____________	40	____________
16	____________	41	____________
17	____________	42	____________
18	____________	43	____________
19	____________	44	____________
20	____________	45	____________
21	____________	46	____________
22	____________	47	____________
23	____________	48	____________
24	____________	49	____________
25	____________	50	____________

NAME ______________________ TEST # __________

51 ____	76 ____
52 ____	77 ____
53 ____	78 ____
54 ____	79 ____
55 ____	80 ____
56 ____	81 ____
57 ____	82 ____
58 ____	83 ____
59 ____	84 ____
60 ____	85 ____
61 ____	86 ____
62 ____	87 ____
63 ____	88 ____
64 ____	89 ____
65 ____	90 ____
66 ____	91 ____
67 ____	92 ____
68 ____	93 ____
69 ____	94 ____
70 ____	95 ____
71 ____	96 ____
72 ____	97 ____
73 ____	98 ____
74 ____	99 ____
75 ____	100 ____

The 3 kilohm load consisted of six 300 ohm headsets each connected across a 500 ohm L-pad. The L-pads were connected in series, thus enabling each listener to adjust volume and still present a constant 3000 ohm load to the circuit.

10. The table below shows further details of the articulation testing described in section nine.

Test #	Listener Scores A	B	C	D	E	Ave.	C db.	db.	Word Lists Used
1*	67	63	70	60	59	64	0	24	1,2
2*	42	35	40	36	37	38	0	18	7,9
3	97	95	91	95	89	93	12	36	20,21
4	86	88	83	85	72	85	12	18	24,25
5	69	61	67	65	53	63	12	12	30,32
6	51	54	49	52	47	51	12	6	12,13
7	46	38	35	42	37	40	12	3	11,14
8	94	91	90	92	85	90	24	30	4,7
9	87	87	89	95	81	88	24	18	8,9
10	80	80	79	66	75	76	24	12	11,13
11	65	55	76	67	58	64	24	6	12,14
12	48	54	59	51	42	51	24	3	22,24

*These tests were used as training tests.

APPENDIX III

THE MANN-WHITNEY U TEST (23, 14)

1. Description.

The Mann-Whitney U test is used to determine whether two independent sets of samples have been drawn from the same population or not. The null hypothesis, H_o, is that the two sets have the same distribution. The alternative hypothesis, H_1, is that one set is stochastically larger or smaller than the other. We accept H_1 if the probability that one single score from one set is larger or smaller than the other is not 1/2.

2. Method.

Call one set of scores X with scores $x_1, x_2, \ldots., x_m$, and the other set Y with scores $y_1, y_2, \ldots., y_n$. First, the two sets of scores are combined and the order statistic formed. Then one set, X or Y, is chosen to form the parameter U. The value of U is given by the number of times that a score in the set, say X, follows a score from Y. A table is consulted giving for each set of m and n the probability that $U \leq U_o$, the value found, if H_o is true. A significance level, α, is chosen. If the value found from the U-test table is greater than α then we say that the sets X and Y came from the same population, or that H_o is true. Conversely, if this value is less than the α chosen, then we say that H_1 is true or that X and Y are from different populations. If the $U = U_o$ calculated is greater than $\frac{mn}{2}$, then we use $U_1 = mn - U_o$ as the value of U for the table.

3. Examples.

Choose $\alpha = 0.01$. This means that it is desired that the values of U should be so small that the probability of their occurrence under H_o is

less than or equal to 0.01.

Tests 12, 24 of section eight:

X = test 12 scores = 36,43,21,40,53,37.

$m = n = 6$

Y = test 24 scores = 58,51,49,56,54,35.

Order statistic:	21,	35,	36,	37,	40,	43,	49,	51,	53,	54,	56,	58.
Set each belongs to:	X	Y	X	X	X	X	Y	Y	X	Y	Y	Y

To find U_o use set X: $U_o = 0 + 1 + 1 + 1 + 1 + 3 = 7$

Table J on page 271 of Siegel gives $P(U \leq 7/H_o) = 0.092$.

This is greater than α so H_1 is rejected.

Tests 7, 19 of section eight:

X = test 7 = 27,30,28,33,28,34.

$m = n = 6$.

Y = test 19 = 49,64,40,44,51,49.

Order statistic:	27,	28,	28,	30,	33,	34,	40,	44,	49,	49,	51,	64.
Set each belongs to:	X	X	X	X	X	X	Y	Y	Y	Y	Y	Y

$U_o = 0 + 0 + 0 + 0 + 0 + 0 = 0$

From the table $P(U \leq 0/H_o) = 0.002$, which is less than α, so H_o is rejected.

4. Significance level.

A significance level of 0.01 was chosen since it was felt that one could not be too rigorous considering the relatively unsophisticated method of testing and the size of the samples. In the study conducted at Montana State College (27) a significance level of 0.001 was used. Lindgren suggests levels of from 0.05 to 0.1, while Siegel uses levels from 0.001 to 0.14. Siegel, in discussing significance gives 0.01 and 0.05 as common values for this type of data.

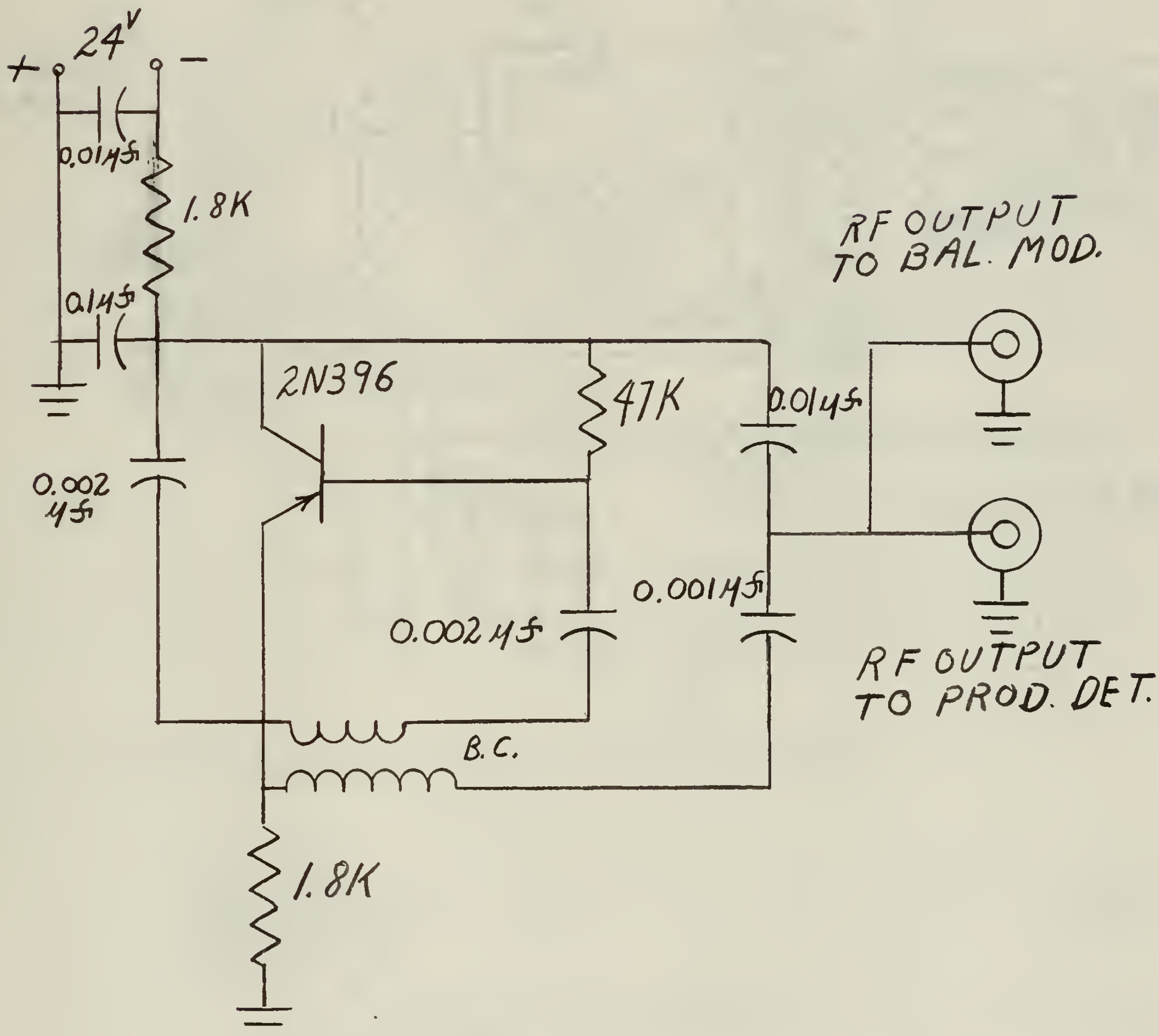
5. Efficiency. The efficiency of this test is quoted by both Lindgren and Siegel to be 0.96 asymptotically, and Lindgren quotes Hodges and Lehmann

as showing that it is always at least 0.864, thus making it one of the most powerful of such tests.

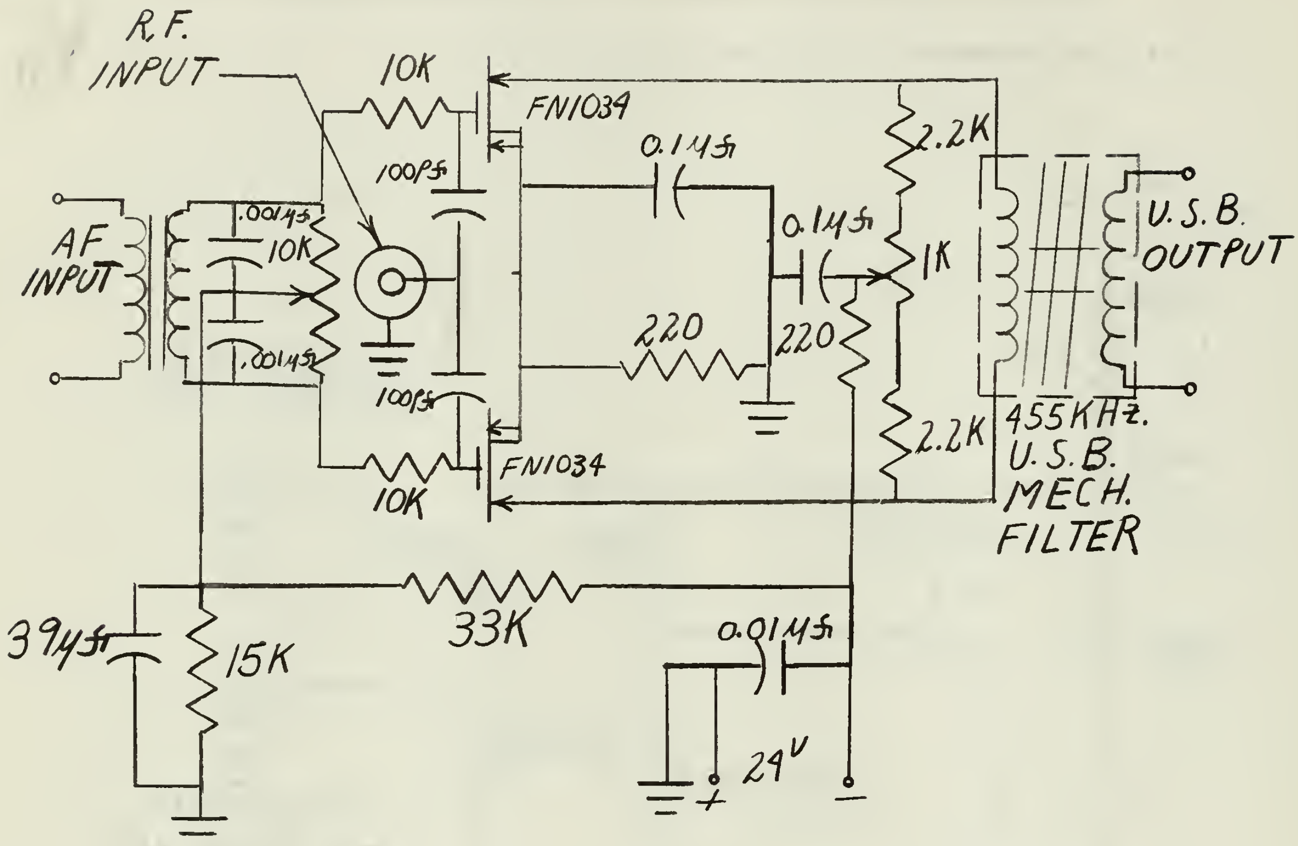
APPENDIX IV

DETAILED SCHEMATIC DIAGRAMS OF R-F SPEECH PROCESSER

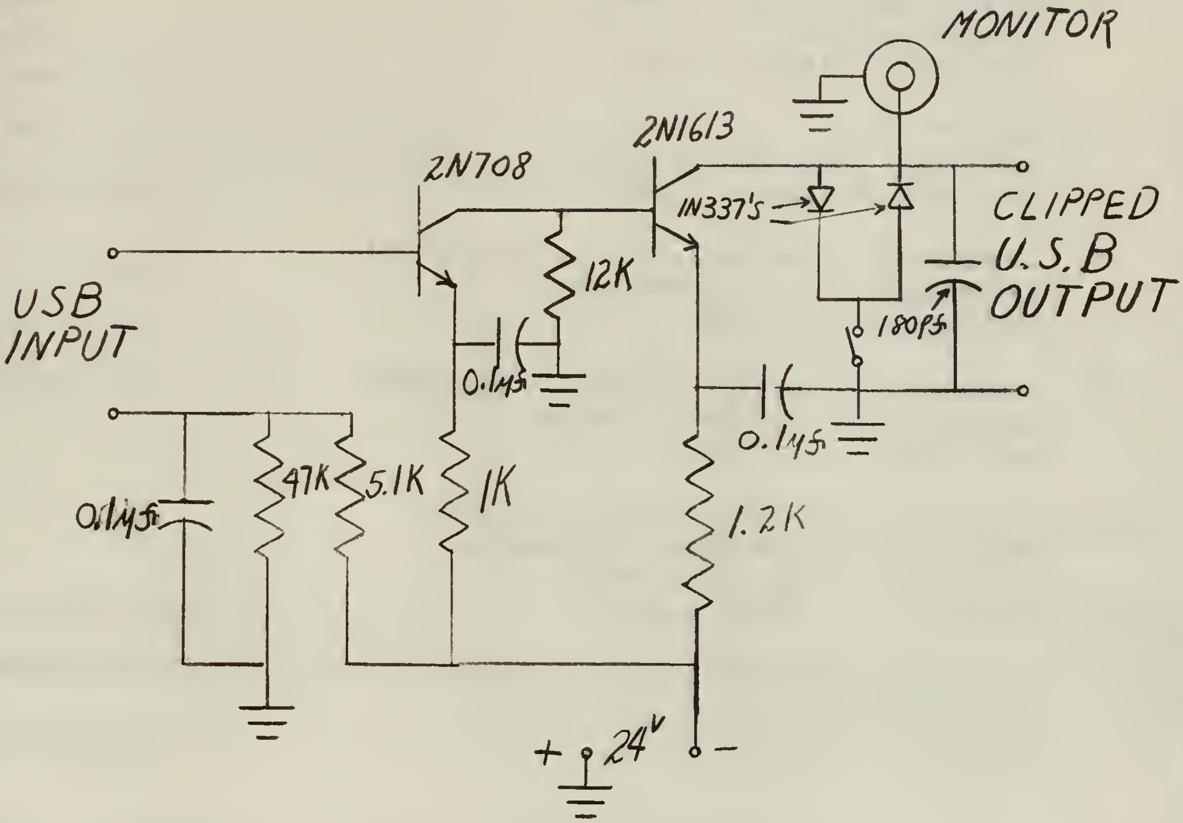
1. Detailed schematic of 455 KHz. L-C oscillator:



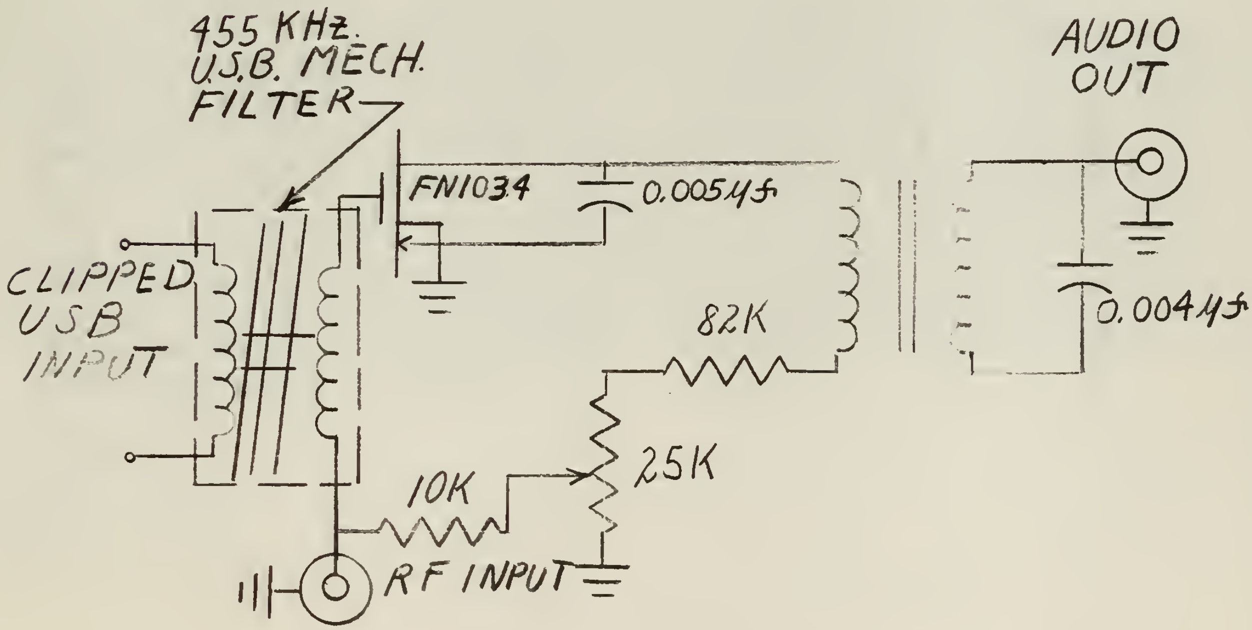
2. Detailed schematic of balanced modulator:



3. Detailed schematic of 455 KHz. amplifier and clipper:



4. Detailed schematic of filter and product detector:



INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	20
2.	Library U.S. Naval Postgraduate School Monterey, California 93940	2
3.	Commandant of the Marine Corps (Code A04C) Headquarters, U.S. Marine Corps Washington, D.C. 20380	1
4.	Commandant of the Marine Corps (Code A03C) Headquarters, U.S. Marine Corps Washington, D.C. 20380	1
5.	Prof. Gerald D. Ewing (Thesis Advisor) Department of Electrical Engineering U.S. Naval Postgraduate School Monterey, California 93940	3
6.	Captain Norman Walter Huddy Jr., USMC 1240 Spruance Road Monterey, California	1

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*

U.S. Naval Postgraduate School
Monterey, California 93940

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP
Not applicable

3. REPORT TITLE
AN INVESTIGATION OF METHODS OF IMPROVING THE INTELLIGIBILITY OF AUDIO FREQUENCY SPEECH IN NOISE

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Master's Thesis in Engineering Electronics

5. AUTHOR(S) *(Last name, first name, initial)*

HUDDY, Norman W.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October 1966	61	30

8a. CONTRACT OR GRANT NO.

b. PROJECT NO.

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. AVAILABILITY/LIMITATION NOTICES

Mennelsen 8/14/69

[redacted]

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Commandant of the Marine Corps (AO4C)
Headquarters, U.S. Marine Corps
Washington, D.C. 20380

13. ABSTRACT

A discussion of the nature of speech is presented, followed by a review of speech processing to date, with emphasis on the characteristics os speech which must be retained for intelligibility. Methods of measuring speech intelligibility are described. The relative merits of abrupt and gradual audio clipping of speech are investigated, and two tone and articulation test results are presented showing that there is no significant difference in these methods of clipping with respect to speech intelligibility. Processing of speech to radio frequencies, filtering and retranslation to audio to improve the peak to average value ratio of the audio frequency prior to transmitting it through a noisy channel is investigated. Two tone and articulation test results are presented showing that this processing results in a 20% improvement in speech intelligibility over audio clipping and filtering.

DD FORM 1 JAN 64 1473

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech						
Speech Processing						
Clipping						
Speech Intelligibility						
Articulation Testing						
Speech Distortion						
R-F Speech Clipping and Filtering						
Gradual Clipping						
Abrupt Clipping						
Audio Speech Clipping and Filtering						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
_______________________________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
_______________________________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
_______________________________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, *or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

DD FORM 1 JAN 64 1473 (BACK)

[redacted]